



181 LDEAKQAMWTSPDQWQPLRYVNTVLQDPVVELFTQAQLDALDGLLYPMIYGAFVDDIAL 240  
241 EGGSSAVAMLTAFMPPEWHTESNRWDVAVYKTMASDDNRALLARWTRDWSRAEAAIAPV 300  
241 NGGSSAVAMLTIPMPWHDDESRWDVAVYKTMATESDNKALLIHWLRTWEDQAAASALLPV 300  
301 AARALQDAGRAALDEVREQFHARAARLGI 329  
301 APMALAECHDIALEEVROOLRAALRRPGL 329

RESULT	2	
W98971		
ID	W98971	W98971 standard; Protein; 331 AA.
AC	W98971;	
DT	10-MAY-1999	(first entry)
DE	Alcaligenes sp.	Protein PoxB.
KW	Alcaligenes; PoxR; PoxA; PoxB; PoxC; PoxD; PoxE; PoxF; PoxG; PoxH;	
KK	PoxI; aromatic; oxidative; petroleum purification; chemical industry;	
OS	drug.	
OS	Alcaligenes sp.	
PN	J11042088-A.	
PD	16-FEB-1999.	
PE	25-JUL-1997;	200625.
PR	25-JUL-1997;	JP-200625.
PA	(TOFU ) TONEN CORP.	
WPI;	99-197820/17.	
DR	N-PSDB; X18867.	
DR	New encoding an aromatic cpd. oxidative decompn. enzyme - useful in	
PT	the fields of petroleum purification, chemical and drug industries	
PT	Claim 1; Page 7-19; 35pp; Japanese.	
PS	The present sequence encodes Alcaligenes sp. protein PoxB. The present	
CC	invention also describes PoxA, PoxK, PoxC, PoxD, PoxE, PoxF, PoxG, PoxH, PoxI,	
CC	and PoxL. The proteins are useful in the fields of petroleum	
CC	purification, chemical industry and drug industry related to the	
CC	synthesis, conversion and decomposition of aromatic compounds.	
CC	Sequence 331 AA;	
SO		

Query Match	48.3%;	Score	832;	DB 1;	Length	331;			
Best Local Similarity	50.0%;	Pred. NO.	3e-81;						
Matches	166;	Conservative	49;	Mismatches	113;	Indels	4;	Gaps	
Qy	1	MTELEKTVDIKPLRHTFAHVAQIGNIGDKTATRYQEGMGAQPOENFHFYRPTWDPDIYFD	60						
Db	1	MOVDTKIQIQPLRQYIGHVARF-GDKPASKRIQATYDVQSEVNFHFYRPTWAPQFIYD	59						
Qy	61	PSRSARMANWYALKDPPOFYVYASWATTARRQODAMESNFEEVERRMIGLRRDDVAARA	120						
Db	60	KRTALEMADWYALKDPPOYYGAYVGTGRCQOEAENFAFVEKRGILLQALPLEWRRL	119						
Qy	121	LQVLVPLRHAANGANNNAQICALGYGVTFVTAPAMFHAMDNIGVQAYLTRILAM--AEP	178						
Db	120	TDGLLPLRHVEAANNNFYCADYGNGTATQACTYCAMDRIGIAQISKIGILLDNGTG	179						
Qy	179	DVLEAKAKWTRDAANOPLRRYVEDTLVADPVEFLFIAGNALDGLLYPLVYDFVDERI	238						
Db	180	AALERARAAMLESAWQPLRRFVEHSFVIEDQFTQVTQNLALDGLLYPLVY-QHADA	238						
Qy	239	ALEGGSAVAMTAFMPEWHETESNRWTDVAVKTMAAESDDNRALLAEWTRDWSARAFAALA	298						
Db	239	VRACGTGLAVVTFEMDWRDDEHVWVDVAVQTTAAASEANRALLISWAGEARAQAFAELR	298						
Qy	299	PVAARALQDAGRALDVEVRFQFHARAARLGIA	330						
Db	299	DVAATTCGCEGRTATCTCFQFQFABIAKIGVA	330						

RESULT	3
R66213	
ID	R66213 standard; Protein; 342 AA.
AC	R66213;
DT	04-AUG-1995 (first entry)

Nocardia corallina alkene monooxygenase subunit-1.  
alkene monooxygenase; subunit-1; epoxidation; amoA gene.  
Nocardia corallina.  
J06292571-A.  
21-OCT-1994.  
06-APR-1993; 105171.  
06-APR-1993; JP-105171.  
(NIHA ) JAPAN ENERGY CORP.  
WPI; 95-009069/02.  
N-PSDB; 079569.  
Alkene monooxygenase and corresp. gene - useful for the  
epoxidation of an alkene  
Claim 1; Page 2; 30pp; Japanese.  
E.coli transformed with the DNA sequence 079569 are able  
catalyse the epoxidation of alkenes. The DNA is derived  
Nocardia corallina and comprises 4 open reading frames.  
and amoc encode subunits 1 and 2 of the alkene monooxygenase  
and ORF amoc encodes a reductase capable of transferring  
from NADH coenzyme to a monooxygenase.  
Sequence 342 AA;  
SO

```

Query Match      10.4%; Score 179.5; DB 1; Length 342;
Best Local Similarity 22.9%; Pred. No. 3.6e-11;
Matches 77; Conservative 55; Mismatches 145; Indels 59; Gaps 14;

31 TRYQEGMMGAQ--PQENFY-----RPTWDPYEIFDPSRAIRMANNYALKDPKQ 79
   || : || || || || || || || || || || || || || || || || ||
29 TEYELIYVGQOSTPEWLHWDPURFDDGRAPWEE-----SSAVRTSEWSAYRDPHQ 81
   || : || || || || || || || || || || || || || || || ||
80 FYASWATTRARQODAMESFEFVESRRMTGLMRDDVAARALDVLPL-----RHAA 131
   || : || || || || || || || || || || || || || || || ||
82 LWQRPYSTCNDQOAL-----ARLPVLTMGSA-----ITPIWSQKILARSAA 127
   || : || || || || || || || || || || || || || || || ||
132 W-----GAMNNAQTCALGYGTVTAPAMTHAMDNLCVAGYLT-RLALAMAPDVLAA 184
   || : || || || || || || || || || || || || || || || ||
128 WPFVEYGLFLSIYAVRQMSDVTQFSWVFQVDMRLQDIYVHLDHQESPEFSAGA 187
   || : || || || || || || || || || || || || || || || ||
185 KATWTRDAWQPLRYVEDTLVADPVELFIAQNLAIDGLLYPLVYDFVDSRIALEG-G 243
   || : || || || || || || || || || || || || || || || ||
188 RPAMWSDSLTPITREVIERIAASQDWVEILLVAGTLVPEPLVGLHAKAEFLSRAPWFGC 247
   || : || || || || || || || || || || || || || || || ||
244 SAVAMLTAFPE--WHITESNWIDAVVKTMAAE---SDDNRALLIARTRDWSARAEAA-- 296
   || : || || || || || || || || || || || || || || || ||
248 TPAVLVALIDSGRHLES-----VQALVRLVCQDPVHGQDQQAIVRWIEEWPCPKAAQA 304
   || : || || || || || || || || || || || || || || || ||
297 -LAPVAARALQDAGRAALDEVREQFHARAARLGTAL 331
   || : || || || || || || || || || || || || || || || ||
295 STPTFESGCTDAKFSANAI SPATANORAAVFEAGAI 340
   || : || || || || || || || || || || || || || || || ||

```

RESULT	4
R81469	R81469 standard; Protein; 343 AA.
ID	R81469; AC
DT	07-AUG-1996 (first entry)
DE	Nocardia corallina alkene mono-oxygenase gene product, amoA.
KW	Alkene mono-oxygenase; indole; indigo production; biosynthesis;
OS	microbial oxidation; dye.
PN	Nocardia corallina B-276.
PD	J08023988-A.
PF	30-JAN-1996.
PP	08-JUL-1994; 179688.
PR	08-JUL-1994; JP-179688.
PS	(NIHA ) JAPAN ENERGY CORP.
PT	WPI; 96-133426/14.
PT	N-PSDB; T17418.
PT	Prepn. of indigo by a microbiological method - by culturing a
PT	microbe having alkene monooxygenase activity to oxidise indole to
PT	indigo
PT	Claim 5; Page 6-7; l1pp; Japanese.
CC	R81469-R81472 are protein products of the Nocardia corallina strain
CC	B-276 alkene mono-oxygenase gene. The gene encodes 4 protein products



Db	172	TWSQDIHLAADNPVSHSNWYMLHFYSGTHGQFLDRIRITYANKGAAI-----FVETGCT	226
QY	259	-----ESNWKIDAVVKTMAAESDDNRLAIRWTROWS-----ARAEAAALAPVAA	302
Db	227	SDASGGNGPYFPQSKWIDFL-----NARKISWV-NWSLADKVETSALMPGAS	274
QY	303	-----RALQDAGRAALDEVRE	318
Db	275	PTGGWTDQAQLSESGKWRQDIQO	297

RESULT 7

R07315 ID R07315 standard; protein; 359 AA.

AC R07315; AC

DT 31-JAN-1991 (first entry)

DE Beta-1,3-glucanase plant pathogenesis-related protein.

DE transgenic plants; disease resistance; chimeric DNA;

KW plant pathogenesis-related protein.

OS synthetic.

PN EP-392225-A.

PD 17-OCT-1990.

PF 21-MAR-1990; 105336.

PR 24-MAR-1989; US-329018.

PR 26-JUN-1989; US-368672.

PR 20-OCT-1989; US-425504.

PR (CIBA ) CIBA GEIGY AG.

PA Ryals JA, Alexander DC, Goodman RM, Meins F, Payne GB;

PI Scinson JR, Neuhaus J-M, Moyer MB;

DR WPI; 90-313983/42.

DR N-PSDB; Q06208.

PT Disease-resistant transgenic plants - obtd. using encoding an

PT inducible pathogenesis-related protein from infected plants.

PS example 32; page 46; 77pp; English.

CC This is the sequence of the plant pathogenesis-related protein

CC (PRP), beta-1,3-glucanase. It confers systemic acquired resis-

CC tance to plants. The corresp. DNA is used, in a chimeric cons-

CC truct, to produce transgenic plant cells or -tissues with the

CC ability to regenerate into plants which are disease resistant.

CC See also Q06179-86, Q06199-Q06207 and Q06829.

SO Sequence 359 AA;

Query Match	5.0%;	Score 85.5;	DB 1;	Length 359;
Best Local Similarity	22.3%;	Pred. No. 0.48;		
Matches	54;	Conservative 34;	Mismatches 71;	Indels 83;
Gaps				

  

QY	58	IFDPSRAIRMANWYALKDPQFYVYASWATTRAQODAMESFEFVESRMIGLMRDVA	117
	:		
	:		
Db	55	LYDPNHGALQ-----ALKG-----SNTFV-----MLGLPNSDK	83
QY	118	ARALDVLVPLRHAAGANMN-----NAQICALG-----YGVTFAPAMFHAM	159
	:		:
	:		:
Db	84	HSASG----MEHARWWQKNVKDFWPDVKIKYIavgNEISPTGTSYLTSFLTPAMVNIY	139
QY	160	DNLCVAGYILRLANREPVL-----EAKATWTRDAW--QPLRRYVEDT-----LV--	206
	:		:
	:		:
Db	140	KATGEAGLGNHNIKVSVDMTLIGNSYPPSGSFRNDARWFTDPIYGFRLDTRAPILLVI	199
QY	207	-----VADPVEFLAQNAL-DG-LIYPLVYDFVDE-KIALE--GGSAVAM	249
	:		:
	:		:
Db	200	YPFYSYSGNPQGISLFXSLFTAFNVVQDGSRSQRYRNLFDAMLDSVYAALERSGGASGVIG	259
QY	250	TA	251
	:		
Db	260	VS	261

  

RESULT	8
W27608	
ID	W27608 standard; Protein; 337 AA.
AC	W27608;









us-09-430-029-3.rag

Wed Sep 27 17:16:53 2000

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 19:19:06 ; Search time 60.65 Seconds  
(without alignments)  
83.660 Million cell updates/sec

Title: US-09-430-029-3  
Perfect score: 1723  
Sequence: 1 MTELEKTDIKPLRHFAHV.....ALDEVREQFHARAARLGIAL 331

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1282	74.4	331	1	US-08-319-387-2
2	186	10.8	327	5	5171684-6
3	178.5	10.4	342	1	US-08-499-215-2
4	87	5.0	560	2	US-08-756-317-8
5	86	5.0	551	2	US-09-033-537A-1
6	84.5	4.9	359	1	US-08-181-271A-98
7	84.5	4.9	359	1	US-08-449-315-98
8	84.5	4.9	359	1	US-08-444-803-98
9	84.5	4.9	359	1	US-08-449-043-98
10	84.5	4.9	359	1	US-08-456-265A-98
11	84.5	4.9	359	1	US-08-455-416-98
12	84.5	4.9	359	1	US-08-455-244-98
13	84.5	4.9	359	2	US-08-454-876-98
14	84.5	4.9	359	2	US-08-457-364-98
15	84.5	4.9	359	2	US-08-456-262-98
16	84.5	4.9	359	2	US-08-456-240-98
17	84.5	4.9	359	2	US-08-455-736-98
18	84.5	4.9	359	2	US-08-971-217-98
19	84	4.9	15281	2	US-08-471-119A-2
20	82	4.8	437	2	US-08-620-605D-4
21	82	4.8	464	2	US-09-005-232A-4
22	79.5	4.6	1810	4	PCT-US95-11684-4
23	78.5	4.6	516	2	US-09-019-201A-3
24	78	4.5	397	2	US-08-282-197C-55
25	78	4.5	457	1	US-08-206-006-2
26	78	4.5	1611	2	US-08-804-227C-5
27	77	4.5	694	2	US-08-895-522-3
28	77	4.5	694	3	US-09-195-391-3

29	76.5	4.4	359	1	US-08-047-413-15	Sequence 15, Appl
30	76.5	4.4	359	3	US-08-229-050-15	Sequence 15, Appl
31	76.5	4.4	3724	2	US-08-804-227C-10	Sequence 10, Appl
32	76.5	4.4	3724	2	US-08-804-198-4	Sequence 4, Appl
33	76	4.4	471	2	US-08-466-583-2	Sequence 2, Appl
34	76	4.4	471	4	PCT-US95-07820-2	Sequence 2, Appl
35	75.5	4.4	4472	2	US-08-804-227C-2	Sequence 2, Appl
36	75	4.4	559	2	US-08-756-317-10	Sequence 10, Appl
37	74.5	4.3	412	2	US-08-851-088-12	Sequence 12, Appl
38	74	4.3	747	2	US-08-895-522-1	Sequence 1, Appl
39	74	4.3	747	3	US-09-195-391-1	Sequence 1, Appl
40	73.5	4.3	262	2	US-08-602-359A-38	Sequence 38, Appl
41	73.5	4.3	269	1	US-07-706-691G-50	Sequence 50, Appl
42	73.5	4.3	269	1	US-08-254-021-50	Sequence 50, Appl
43	73.5	4.3	269	2	US-08-618-446-50	Sequence 50, Appl
44	73.5	4.3	4545	2	US-08-804-227C-14	Sequence 14, Appl
45	73	4.2	268	2	US-08-835-099A-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1  
US-08-319-387-2  
; Sequence 2, Application US/08319387  
; Patent No. 5543317  
; GENERAL INFORMATION:  
; APPLICANT: Shields, Malcolm S.  
; APPLICANT: Francesconi, Stephen C.  
; TITLE OF INVENTION: Microbial Degradation of Trichloroethylene,  
; Dichloroethylenes and Aromatic Pollutants  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/319,387  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/167,457  
; FILING DATE: 15-DEC-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/694,718  
; FILING DATE: 02-MAY-1991  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Whitlock, Ted W.  
; REGISTRATION NUMBER: 36,965  
; REFERENCE/DOCKET NUMBER: UWF-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 331 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-319-387-2

Query Match 74.4%; Score 1282; DB 1; Length 331;  
Best Local Similarity 72.9%; Pred. No. 2.9e-132;  
Matches 240; Conservative 32; Mismatches 57; Indels 0; Gaps 0;

QY 1 MTIELKTVDIKPLRHTFAHVAQNTGGDKTATRYQEGMGAQPOENFYRPTWDPDYELFD 60  
DB 1 MTIDLKTRKPLRHTTHVAQYICADKAASRYEGTVGAQPAANFYRPTWDPHEHLEFD 60  
QY 61 PPSAIRMANWYALKDRPQFYASWATTARQQDAMESNEFEVSRMIGLMDRDVAA 120  
DB 61 TSRTAIQMKOMYALKDRPQFYASWATTARQQDAMESNEFEVSRMIGLMDRDVAA 120  
QY 121 LDVPLRHAAGNNAQICALGYCTVFTAPAFHAMDNLGVAQYILTRIALAMAEPPDV 180  
DB 121 LSVLPLRHAAGNNAQICALGYCTVFTAPAFHAMDNLGVAQYILTRIALAMAEPPDV 180  
QY 181 LEAAKATWTRDAWQPLRRYVEDTLVADPVLEFTAQNALDGLLYPLVYDFRVDRIAL 240  
DB 181 LDEAKQAWMTSPDWQPLRRYVENTLVQDPVLEFTAQNALDGLLYPMIYGAFVDDYIAL 240  
QY 241 EGGSAVAMLTAFMEPWHITESNRWIDAVVYKMAESDDNRALLARWTRDWSARAEALAPV 300  
DB 241 NGSASAVAMLTAFMEPWHITESNRWIDAVVYKMAESDDNRALLARWTRDWSARAEALAPV 300  
QY 301 AARALQDAGRAALDEVEQEFAHARALGI 329  
DB 301 AEMALAEHGDALDEVRQQLRAALRRPCL 329

## RESULT 2

5171684-6  
Patent No. 5171684  
APPLICANT: YEN, KWANG-MU; BLATT, LAWRENCE M.; KARL, MICHAEL R.  
TITLE OF INVENTION: BIOCONVERSIONS CATALYZED BY THE TOULUENE  
MONOOXYGENASE OF PSEUDOMONAS MENDOCINA KR-1  
NUMBER OF SEQUENCES: 41  
CURRENT APPLICATION DATA:  
FILING DATE: 05/07/1990, 374  
FILING DATE: 28-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 177,631  
FILING DATE: 05-APR-1988

SEQ ID NO:6:

LENGTH: 327

5171684-6

Query Match 10.8%; Score 186; DB 5; Length 327;  
Best Local Similarity 22.7%; Pred. No. 2.5e-12;  
Matches 73; Conservative 61; Mismatches 136; Indels 52; Gaps 12;

QY 1 MTIELKTVDIKPLRHTFAHVAQNTGGDKTATRYQEGMGAQPOE-----NFHYRPTWDP 54  
DB 1 MSFESK-----KPMR-TWSHLAE-----MRKAPSEYDIVSRKLHY-STNNP 39  
QY 55 D--YEIEDPS-----RSAIRMANWYALKDRPQFYASWATTARQQDAMESNFE 101  
DB 40 DSPWELSPDSPMNLWYKQYRNASPLKHDNDWDAFTDPQLVYRTYTNLMQDQGESYVQSLEFD 99  
QY 102 FVSRMRIGLMDRDVAAALDVLPLRHAAGNNAQICALGYCTVFTAPAFHAMDNL 161  
DB 100 QNFREFHQDMVREGWEHTMARCYSPLRYLFHCLQMSAYVQQAAPASTISNCCILQATDS 159  
QY 162 LGVAQYILTRIA-----LAMAEPD--VLEAAKATWTRDAWQPLRRYVEDTLVADPVLEF 214  
DB 160 L---RWLTHYATRYHELSLTTPDAGLGEHERELMEKEPGWGLRMEKQLTAFDWGEAF 216  
QY 215 IAQNALDGLLYPLVYDFRVDRIALGEGSAVAMLTAFMEPWHITESNRWIDAVVYKMAE 274  
DB 217 VSLNLVVKPMIVESIF-KPLQQQAWENNNDTLILLIDSQKDAERHSWSKALVK-HALE 274  
QY 275 SDDNRALLARWTRDWSARAEAA 296

: || ||: | | | | |

DB 275 NPDNHAVIEGWIEKWRPLADRA 296

## RESULT 3

US-08-499-215-2  
Sequence 2, Application US/08499215  
Patent No. 5612204  
GENERAL INFORMATION:  
APPLICANT: Sasaki, Hisashi  
APPLICANT: Miura, Akira  
TITLE OF INVENTION: BIOLOGICAL DEGRADATIVE  
INVENTION: TREATMENT OF CHLORINE-SUBSTITUTED ETHYLENE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
STREET: 2100 Pennsylvania Avenue  
CITY: N.W.  
STATE: Washington D.C.  
COUNTRY: U.S.A.  
ZIP: 20037-3202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
SOFTWARE: & WordPerfect version 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08499,215

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP Hei-6-179689

FILING DATE: 08-JUL-1994

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 342 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-499-215-2

Query Match 10.4%; Score 178.5; DB 1; Length 342;

Best Local Similarity 22.9%; Pred. No. 1.8e-11;

Matches 77; Conservative 55; Mismatches 145; Indels 59; Gaps 14;

QY 31 TRYQEGMGAQ--POENFY-----RPTWDPDYELFDPSSAIRMANWYALKDRPQ 79

DB 29 TEYELYTVGQOSTPDWELHVDWPLRFDDGRAPWEE-----SSAVTSEWSAYRDPHQ 81

QY 80 FYASWATTARQQDAMESNEFEVSRMIGLMDRDVAAALDVLVPL-----RHAA 131

DB 82 LWQRPYVSTCQDQAL-----ARLPVLTMSRA-----TTPWSOKILARSYAA 127

QY 132 W-----GAMNNAQICALGYCTVFTAPAFHAMDNLGVAQYLT-RLALAMAEPPDVEA-A 184

DB 128 WPFVEYGLFLSLAYAVRQAMSDTVQSVFQAVDRMRLIQDIVHLDLQESPEESDAGA 187

QY 185 KATWTRDAWQPLRRYVEDTLVADPVLEFTAQNALDGLLYPLVYDFRVDRIALG-G 243

DB 188 REAMMSDTLPIREVIEWERIASAQDWVEILVAGTLVFEPVLVGHILAKAEILFRRAPDFGDG 247

QY 244 SAVAMLTAFME--WHTESNRWIDAVVYKMAE-----SDDNRALLARWTRDWSARAEAA-- 296

DB 248 TPPAVIASALLDSGRHLES---VQALVRLVCQDPVHGQNOATVRRWTEEQPRCKAAQ 304

QY 297 -LAPVAARALQDAGRAALDEVEQEFAHARALGIAL 331

DB 305 SFLPTFSCGDIAKESANALSALANQANRAAVEGAGI 340

## RESULT 4

US-08-756-317-8

```

: Sequence 8, Application US/08756317
: Patent No. 5849894
: GENERAL INFORMATION:
: APPLICANT: Clemente, Thomas E.
: APPLICANT: Kishore, Ganesh M.
: APPLICANT: Mitsky, Timothy A.
: APPLICANT: Stark, David M.
: TITLE OF INVENTION: Improved Rhodospirillum Rubrum
: TITLE OF INVENTION: Poly-B-Hydroxyalkanoate Synthase
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: TX
: COUNTRY: USA
: ZIP: 77210-4433
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08756,317
: FILING DATE: 25-NOV-1996
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/007,693
: FILING DATE: 29-NOV-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Patterson, Melinda L.
: REGISTRATION NUMBER: 33,062
: REFERENCE/DOCKET NUMBER: MOBT:008
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (713) 787-1400
: TELEFAX: (713) 787-1440
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 560 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
:
: S-08-756-317-8

```

Query Match 5.0%; Score 87; DB 2; Length 560;  
Best Local Similarity 21.8%; Pred. No. 0.39;  
Matches 75; Conservative 43; Mismatches 120; Indels 106; Gaps 21;

```

QY 5 LKTVITKPLRHTFAHQAQNTGGDKTATRYOEGMGAQPOFNHYRTPWDPDYEIFD--ps 62
Db 201 LELIQKPL-----GER---QYAKPLLIYPPQINKYY-----IFDLSPE 236
QY 63 RSAIRANWYALKDPQFYFASHATTAAQQDAMENSFEEVSRRMIGLUMRDDVAARALD 122
Db 237 KSFVQ---YALKNNLQVFVISH---RNPDA-----QHREWGL---STVEALD 275
QY 123 VLPLRHAAWGA-MNNAQICALGYGTVTAPAMFH---AMDNLGVAQYLTRL-----171
Db 276 QATEVSEHETGSRVSNLAGACA---GGLTVAALLGLHQLVRRLKRVSSVTYLLSLDSQM 332
QY 172 ---ALAMAEEDVLEAAKA-----TWTR--DAAWQPLRRYVEDTLVVAD 209
Db 333 ESPAMLFADQTLLESSKRRSYQHGVLDGRDMAKVFAWMPNDLIW---NYWNNVYLLGRQ 389
QY 210 PVLEFIA---QNLALDGLLYPIWYDFVDERI-----ALE-GGSVAAMLTAFPMEWHETS 260
Db 390 PPAFDILYWNNDNTRLPAAFHGGELLDFLKHNPILTRPGALEVSTAVDLGKVAIDSFHVAG 449
QY 261 -----NEWIDAVVYKTHMAASDQNRALLARWTKDWSARAEAAALAP 299
Db 450 ITDHTTPW-DAVYRSALLIGGRRFFILSN-----SGHIQSIINLP 487

```

RESULT 5

```

US-09-033-537A-1
; Sequence 1, Application US/09033537A
; Patent No. 5958083
; GENERAL INFORMATION:
; APPLICANT: Onishi, Masahiro
; APPLICANT: Fich, Mercete
; APPLICANT: Toft, Annette Hanne
; APPLICANT: Sh. lein, Martin
; TITLE OF INVENTION: Prevention of Back-Staining
; TITLE OF INVENTION: In Stone Washing
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958083o No. 5958083disk of No. 5958083th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; COUNTRY: NY
; COUNTRY: U.S.A.
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/033,537A
; FILING DATE: 02-MAR-1998
; CLASSIFICATION: 008
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 0993/95
; FILING DATE: 08-SEP-1995
; APPLICATION NUMBER: PCT/DK96/00364
; FILING DATE: 03-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4492.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 551 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-033-537A-1

```

Query Match 5.0%; Score 86; DB 2; Length 551;  
Best Local Similarity 18.6%; Pred. NO. 0.49;  
Matches 49; Conservative 42; Mismatches 78; Indels 94; Gaps 13;

Qy	132	WGANNNAQICALGYGTVFAPAM-----FHAMDNLGVAOYLTRIALAMAEPPDVEAA	184
		:   :    :	
Dd	53	WGINVFEAAMYTABDGXI-TDPGVKNKVKEAQAQSIDGLGYVIDWHILSDGNPNITYKAQ	111
		:   :    :	
Qy	185	KAIWTRDAA-----WQLRRYYEDTL-----VVADP	210
		:::     :    :	
Dd	112	SKAFFGEATYLGNTPNVIEIANEPNGNVSWADVKSIAEEVITAIRADPDGVVTVGSP	171
Qy	211	-----VELFIAQNALDGLLYPL-----VYDRVDRL--ALEGSAVAMLTAFFPEWHT	258
		:   :     :     :	
Dd	172	TWSODIHLAADNPVSNSVMIALHFTSGHGQFLRKTIYMNKGAAI-----FVTEGT	226
		:   :     :     :	
Qy	259	-----ESNRWIDAVVKIMAESDDNRIALLARWTOWS-----ARAEEAALAPVAA	302
		:   :     :     :	
Dd	227	SDASGGNGPYFPQSKEWIDL-----NARKISWV-NWLSADKVETSAALMPGAS	274
		:   :     :     :	
Qy	303	-----RALQDAGRAALDEVR	318
		:   :     :     :	

Db 275 PTGWTDAQLSESGKWRDQIRQ 297

RESULT 6

US-08-181-271A-98

; Sequence 98, Application US/08181271A

; Patent No. 5614395

GENERAL INFORMATION:

; APPLICANT: Ryals, John A.

; APPLICANT: Alexander, Danny C.

; APPLICANT: Beck, James J.

; APPLICANT: Duesing, John H.

; APPLICANT: Friedrich, Leslie B.

; APPLICANT: Goodman, Robert M.

; APPLICANT: Harms, Christian

; APPLICANT: Meins, Jr., Frederick

; APPLICANT: Montoya, Alice

; APPLICANT: Moyer, Mary B.

; APPLICANT: Neuhaus, Jean-Marc

; APPLICANT: Payne, George B.

; APPLICANT: Sperison, Christoph

; APPLICANT: Stinson, Jeffrey R.

; APPLICANT: Uknes, Scott J.

; APPLICANT: Ward, Eric R.

; APPLICANT: Williams, Shericea C.

TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: New York

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/181,271A

FILING DATE: 13-JAN-94

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/093,301

FILING DATE: 16-JUL-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/937,197

FILING DATE: 6-NOV-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/678,378

FILING DATE: 1-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/305,566

FILING DATE: 6-FEB-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/165,667

FILING DATE: 8-MAR-1988

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/042,847

FILING DATE: 6-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/632,441

FILING DATE: 21-DEC-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/425,504

FILING DATE: 20-OCT 1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/848,506

FILING DATE: 6-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/768,122

FILING DATE: 27-SEP-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/580,431

FILING DATE: 7-SEP-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/368,672

FILING DATE: 20-JUN-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/329,018

FILING DATE: 24-MAR-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/045,957

FILING DATE: 12-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919)541-8614

TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 98:

SEQUENCE CHARACTERISTICS:

LENGTH: 359 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-181-271A-98

Query Match 4.9%; Score 84.5; DB 1; Length 359;

Best Local Similarity 21.9%; Pred. No. 0.37;

Matches 53; Conservative 35; Mismatches 71; Indels 83; Gaps 14;

QY 58 IFDPSRAIRMANWYALKDPQFYVASWATTFARQDAMESNFEVESRMIGLMDOVA 117

Db 55 LYDPNHGALQ-----ALRG-----SNIEV-----MLGLNSDVK 83

QY 118 ARALDVLPLRHAAGANN-----NAQICALG-----YGTFTAPAMFHAM 159

Db 84 ----HIASSMEHARWVQKNVDFWPKIKYIAGVNEISPYTGTSYLTSETLPAMVNIY 139

QY 160 DNLGVAQYLTRLALAMAEFDVL-----BAKATWTRDAW--QPLRRYVEDT----LV-- 206

Db 140 KAIGEAGLGNLIKVSTSDMTLIGNSYPPSGSFRNDARFWFDPVIGFLDTRAPLLVNI 199

QY 207 -----VADPVLFIAQNLAL-DG-LLYPLVYDRFVDE-RIALE--GGSAVAML 249

Db 200 YPFYSYSGNPGQISLPYSILFTAPNVVVQDSRQYRNLFDMLDYSVAALERSGGASGVIV 259

QY 250 TA 251

Db 260 VS 261

RESULT 7

US-08-449-315-98

; Sequence 98, Application US/08449315

; Patent No. 5650505

GENERAL INFORMATION:

; APPLICANT: Ryals, John A.

; APPLICANT: Alexander, Danny C.

; APPLICANT: Beck, James J.

; APPLICANT: Duesing, John H.

; APPLICANT: Friedrich, Leslie B.

; APPLICANT: Goodman, Robert M.

; APPLICANT: Harms, Christian

; APPLICANT: Meins, Jr., Frederick

; APPLICANT: Montoya, Alice

; APPLICANT: Moyer, Mary B.

; APPLICANT: Neuhaus, Jean-Marc

; APPLICANT: Payne, George B.

; APPLICANT: Sperison, Christoph

; APPLICANT: Stinson, Jeffrey R.



APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESS: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US 08/449,315  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmet, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/Pl/CSG 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689

INFORMATION FOR SEQ ID NO: 98:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 359 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-449-315-98  
Query Match 4.9%; Score 84.5; DB 1; Length 359;  
Best local Similarity 21.9%; Pred. No. 0.37;  
Matches 53; Conservative 35; Mismatches 71; Indels 83; Gaps 14;  
QY 58 IFDPSRAIRMANWYALKDPQFYVASWATTRARQODAMESFEFVESRRMIGLMRDDVA 117  
DB 55 LYDPNHALQ-----ALKG-----SNIEV-----MLGLPNSDVK 83  
QY 118 ARALDVLVPLRHAAGANN-----NAQICALG-----YGVFTAPAMFHAM 159  
DB 84 -----HIASGMEHARWVQKNVKDFWPDVKIKYIAGVNEISPVGTISLTSPAMVNTY 139  
QY 160 DNLGVAQYLTRLALAMAEEDVL-----EAKATWTFDAW---QPLRRYVEDT-----LV-- 206  
DB 140 KAIEAGLGNKIKVSTVDMTLIGNSYPPSGSFRNDAEFVDPIVGFLLRTRAPLLVNI 199  
QY 207 -----VADPVELFIAQNLAL-DG-LLYPLVYDFRVE-RIALE--GGSAVAML 249  
DB 200 YPFYSYSGNPGQISLPYSLFTAPNVVQDSRQYRNLFDMLDVYVALERSGGASGVIV 259  
QY 250 TA 251  
DB 260 VS 261  
RESULT 8  
US-08-444-803-98  
Sequence 98, Application US/08444803  
Patent No. 5654414  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08444,803





	Query Match	4.9%; Score 84.5; DB 1; Length 359;
	Best Local Similarity	21.9%; Pred. No. 0.37;
	Matches 53; Conservative 35; Mismatches 71; Indels 83; Gaps 14;	
QY	58 IFDPSSATRMNVAALKDPRQYYASWATTRAQODAMESFEESRMIGLMRDDVA	117
Ddb	55 LYDPNHGALQ-----ALKG-----SNIEV-----MLGIPLNSDVK	83
QY	118 ARALDVLVPLRHAAWGANKMN-----NAQICALG-----YGTVFTAPAFAM	159

```
Db 84 -----HIASGMEHARWVQKVKDFWPDVKIKYIAVGNESPVIGTSYLTSFLTPTAMVNIY 139
QY 160 DNLGVAQYLTRLALAMAEFDVL-----EAAKATWTRDAW--OPLRRYVEDT----IV-- 206
:|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 KAIGEAGLGNKIKVSTVDMTLIGNSYPPSQGSFRNDARWFVDPIVGFRLDTRAPLVNI 199
:|: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 207 -----VADPVELFIAQNLAL-DG-LLYPLVYDRFVDE-RIALE--GGSAVAML 249
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
Db 200 YPFYSYGNPGQISLPSYSLFTAPNVVVQDGSQRQYRNLDFDMLDSVYAALERSGGASVGI 259
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 250 TA 251
Db 260 VS 261

RESULT 12
US-08-455-244-98
; Sequence 98, Application US/08455244
; Patent No. 5789214
; GENERAL INFORMATION:
; APPLICANT: Ryals, John A.
; APPLICANT: Alexander, Danny C.
; APPLICANT: Beck, James J.
; APPLICANT: Duesing, John H.
; APPLICANT: Friedrich, Leslie B.
; APPLICANT: Goodman, Robert M.
; APPLICANT: Harms, Christian
; APPLICANT: Meins, Jr., Frederick
; APPLICANT: Montoya, Alice
; APPLICANT: Moyer, Mary B.
; APPLICANT: Neuhaus, Jean-Marc
; APPLICANT: Payne, George B.
; APPLICANT: Sperison, Christoph
; APPLICANT: Stinson, Jeffrey R.
; APPLICANT: Uknes, Scott J.
; APPLICANT: Ward, Eric R.
; APPLICANT: Williams, Shericea C.
; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: New York
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,244
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/181,271
; FILING DATE: 13-JAN-94
; APPLICATION NUMBER: US 08/093,301
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/937,197
; FILING DATE: 6-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/678,378
; FILING DATE: 1-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/305,566
; FILING DATE: 6-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/165,667
; FILING DATE: 8-MAR-1988
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/042,847
; FILING DATE: 6-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/632,441
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/425,504
; FILING DATE: 20-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,506
; FILING DATE: 6-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/768,122
; FILING DATE: 27-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/580,431
; FILING DATE: 7-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/368,672
; FILING DATE: 20-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/329,018
; FILING DATE: 24-MAR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/045,957
; FILING DATE: 12-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: S-19825/PL/CCG 1727
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919)541-8614
; TELEFAX: (919)541-8689
; INFORMATION FOR SEQ ID NO: 98:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-455-244-98
```

```
Query Match 4.9%; Score 84.5; DB 1; Length 359;
Best Local Similarity 21.9%; Pred. No. 0.37;
Matches 53; Conservative 35; Mismatches 71; Indels 83; Gaps 14;
```

```
QY 58 IFDPSRAIRMANWYALKDPQFYFVYASWATTRARQQDAMESNFVFESRRMIGLMRDVYA 117
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
Db 55 LYDPNHEGALQ-----ALKG-----SNIEV-----MLGLPNSDVK 83
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 118 ARALDVLVPLRHAANGANN-----NAQICALG-----YGVFTAPAMFHAM 159
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
Db 84 ----HIASGMEHARWVQKVKDFWPDVKIKYIAVGNESPVIGTSYLTSFLTPTAMVNIY 139
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 160 DNLGVAQYLTRLALAMAEFDVL-----EAAKATWTRDAW--OPLRRYVEDT----LV-- 206
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
Db 140 KAIGEAGLGNKIKVSTVDMTLIGNSYPPSQGSFRNDARWFVDPIVGFRLDTRAPLVNI 199
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 207 -----VADPVELFIAQNLAL-DG-LLYPLVYDRFVDE-RIALE--GGSAVAML 249
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
Db 200 YPFYSYGNPGQISLPSYSLFTAPNVVVQDGSQRQYRNLDFDMLDSVYAALERSGGASVGI 259
:|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 250 TA 251
Db 260 VS 261
```

```
RESULT 13
US-08-454-876-98
; Sequence 98, Application US/08454876
; Patent No. 5804693
; GENERAL INFORMATION:
```





```

Db      200 YPYFSYSGNFQISLPYSLFTAPVVVQDGSROYRNLFDAMLDSVYAALERSGGASVGIV 259
QY      250 TA 251
        :
Db      260 VS 261

Search completed: September 26, 2000, 20:22:37
Job time: 3811 sec

```

Query Match	4.98;	Score	84.5;	DB	2;	Length	359;
Best Local Similarity	21.98;	Pred.	No. 0.37;				
Matches	53;	Conservative	35;	Mismatches	71;	Indels	83; Gaps
							14;
QY	58	IFDPSRAIRMANVYALKDPQFYASWATTARQQADAMESNFESERMIGLMRDVA	117				
Db	55	LYDENHGALQ-----ALKG-----SNIEV-----MGLPNSDVK	83				
QY	118	ARAULDVLPLRHAAGANNN-----NAQICALG-----YGTVTATPAMFHAM	159				
Db	84	---HIASGMGEHARWVGKNVKDFWPDKIKYIAGNGEISPVGTGYLTSLTAFAMVIY	139				
QY	160	DNLGVAOYLTRLATAEPDVL----EAAKATWTTRDAAW--QLRRRYVEET----	206				
Db	140	KAIAGEALGNKKIYSTVDMTLIGNSPYSGSFNDARFWFDPFLVGELRTDRAPLLVNI	199				
QY	207	-----VADPVELFTIAQNIAL-DG-LLXPLVNYDREVDE-RIALE--GGSAVAML	249				



---

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 19:23:36 ; Search time 82.1 Seconds  
(without alignments)  
249,499 Million cell updates/sec

Title: US-09-430-029-3

Perfect score: 1723

Sequence: 1 MTELEKTVDIKPLRHTFAHV.....ALDEVREQFHARAALRGIAL 331

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_64:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	779.5	45.2	331	1	B37831	phenol 2-monooxyge
2	778.5	45.2	331	1	S44304	phenol 2-monooxyge
3	640.5	37.2	333	1	S47288	phenol 2-monooxyge
4	123	7.1	394	2	S15208	methane monooxygen
5	105.5	6.1	1092	2	T18305	replication factor
6	105.5	6.1	1092	2	T18306	replication factor
7	105	6.1	505	2	S39962	endoglucanase - Er
8	103	6.0	387	2	JL0101	methane monooxygen
9	99	5.7	580	2	G70763	probable maltoolig
10	97.5	5.7	8563	2	T30226	polyketide synthas
11	96	5.6	1238	2	T03465	probable exonuclea
12	91.5	5.3	472	1	G70932	probable monooxygen
13	90.5	5.3	697	2	C75525	molybdopterin oxid
14	90.5	5.3	875	2	G75442	ATP-dependent Clp
15	90	5.2	539	2	A58538	4-hydroxybenzoate-
16	89.5	5.2	415	2	B70898	probable exonuclea
17	89.5	5.2	522	2	A75412	conserved hypothet
18	88.5	5.1	467	2	S58233	PopQ protein - Rhi
19	88.5	5.1	474	2	G72658	probable Vir B11 A
20	88.5	5.1	504	2	S54744	cellulase (EC 3.2.
21	88.5	5.1	509	2	T45408	pet112-like protei
22	88	5.1	1309	2	T00078	probable RNA-direc
23	87.5	5.1	463	2	S01984	hypothetical prote
24	87.5	5.1	1147	2	T35781	hypothetical prote
25	87.5	5.1	1396	2	S36851	L-shaped tail fibe
26	87	5.0	560	2	S29307	poly(3-hydroxyalka
27	87	5.0	573	2	T10037	hypothetical prote
28	87	5.0	1240	2	T04193	hypothetical prote
29	86.5	5.0	454	2	T46818	probable monooxyge

30	86.5	5.0	555	2	F65254	ABC transporter in
31	86.5	5.0	1273	2	E72611	probable Arp-depen
32	86.5	5.0	3158	3	T17483	peptide synthetase
33	85.5	5.0	474	2	JX0301	ubiquinol--cytochr
34	85.5	5.0	509	2	D70856	phosphatase gntB - My
35	85.5	5.0	526	2	A81135	O-antigen biosynth
36	85.5	5.0	1275	2	T18556	glucan endo-1,3-be
37	85	4.9	369	2	S12406	recap protein - Stre
38	85	4.9	456	2	A31844	hydroxyneurosporen
39	85	4.9	485	2	S32171	probable fadE6 pro
40	85	4.9	731	2	B70628	glucan endo-1,3-be
41	84.5	4.9	359	2	A30758	1,3-beta-glucanase
42	84.5	4.9	365	2	S13594	glucan endo-1,3-be
43	84.5	4.9	370	2	A39115	glucan endo-1,3-be
44	84.5	4.9	370	2	B39115	glucan endo-1,3-be
45	84.5	4.9	606	1	S59634	endo-1,4-beta-xyla

ALIGNMENTS

RESULT 1

B37831

phenol 2-monooxygenase (EC 1.14.13.7) chain P1 - Pseudomonas sp. (strain CF600)  
C:Species: Pseudomonas sp.

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: B37831

R:Nordlund, I.; Powlowski, J.; Shingler, V.

J. Bacteriol. 172, 6828-6833, 1990

A:Title: Complete nucleotide sequence and polypeptide analysis of multicomponent phen

A:Reference number: A37831; MUID:91072230

A:Accession: B37831

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-331 <NOR>

A:Cross-references: GB:M60276; GB:M37764; MID:gl51449; PIDN:AAA25940.1; PID:gl51451

C:Superfamily: phenol 2-monooxygenase component L

C:Keywords: oxidoreductase

Query Match 45.2%; Score 779.5; DB 1; Length 331;  
Best Local Similarity 46.8%; Pred. No. 4.7e-59;  
Matches 155; Conservative 54; Mismatches 115; Indels 7; Gaps 4;

QY	1	MTELEKTVDIKPLRHTFAHVAQNI	GDKTATRYQEGMGAQPOENH	HYRETWDPDEIFD	60
DB	1	MSVEIKTNTVDPTRQTYGNLQRRF	-GDKFASRYQEAASYDIEAVTN	PHYRLWDPQHLDH	59
QY	61	PSRSATRMANNWALKDPRFYFAS	WATTRARQODAMESNFEFVES	RRMIGLMRDDVAARA	120
DB	60	PIRTAIRMIDHWKVTDPQFYFAY	QVQTRARMQEATEHAYGFCEK	RELLSRLPAETQAKL	119
QY	121	LDVLVPLRHAANGANNNAQICAL	GYGVTFTAPMFAMNDNLGVAQ	YLTFLALAM--AEP	178
DB	120	LRLCLVPLRHAELGANNNSN	AGDSIAATVOMHIYQAMDRLG	MGYLSRIGLLDGGTG	179
QY	179	DVLEAKAKATWTRDAAMQPLRR	VYVEDTLVADPVELLFAQNLA	LDGLLYLVVYDRFDERI	238
DB	180	EALDQAKATWLDPTIQGLRRY	VEDSFVIRDMFELGAQNLYLD	GLQLPLMYQRF--DQWL	238
QY	239	ALEGGSAVALMTAFMP	EWHTESNRWIDAVVKTMAAES	DDNRALLARWTRGWSARA	298
DB	239	TENGSGSDVAMLI	TEFMRDVGESTRWYDAMFKT	VLAENDANREQVQAWLEW	WEPRATEALL 298
QY	299	PVAARALQDAGRAALDEV	REQFHARAALRGI 329		
DB	299	PLAEEA---TGIAAL	DEVRSFAFATRLQKTGL 326		

RESULT 2

S44304

phenol 2-monooxygenase (EC 1.14.13.7) component L - Pseudomonas putida

N:Alternate names: phenolhydroxylase chain B



Db 47 KPLSEYQSLSCYQNPNDWIAAGLDGWDWTKQFGGRPSWGN-----STELRTTDWY 99  
 QY 73 ALKDP-RQFY--YASWATTRARQDAMESNEFEVESRRMIG-LMRDDVAARALDVLPLR 128  
 Db 100 RHRDPARWHHPYKDKSEARYQRFELAAVSSSGSIRTDIPYRDEILNKYFALLYSE 159  
 QY 129 HAAGANNNAQICALGYGVFTAPAMFHAMDNGLGVAQ--YILRLALAMAEPOV---LEA 183  
 Db 160 YGLFNAHSSVGRDC--LSDFIRQTAVEAALDKVDNAQMIQMERLFTAKLVPGFEDASIDV 216  
 QY 184 AKATWTRDAAQPLRRYVEDTLV-VADPELVFIAQNALDGLLPLVYDVRVDERIALEG 242  
 Db 217 PKKITWTDPIYSGARATVQEIWOGVODWNEILWAGHAVMIATFGQFARREFFRLATVYG 276  
 QY 243 GSAVAMLTAFMPEWHITESNRWIDAV-VKTMAAESD---DNEALLARWTRDWSARAALA 298  
 Db 277 DTLPPFTAQSQTFQTRTGAIIDLFFVCLANDSEFGAHHNRTFLNAWTEHYLASSVAALK 336  
 QY 299 P-VAARALQDAGRAALDEVREQFHARAARLGI 329  
 Db 337 DFVGLYAKVEKSA--DRSRRRLRGAARASAI 366

## RESULT 5

Tl8305  
 replication factor C like protein - Emericella nidulans  
 C:Species: Emericella nidulans, Aspergillus nidulans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: Tl8305  
 R:Kafer, E.; May, G.  
 Gene 191, 155-159, 1997  
 A:Title: The uvsF gene region in Aspergillus nidulans codes for a protein with homology  
 A:Reference number: Z18873  
 A:Accession: Tl8305  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1092 <KAF>  
 A:Cross-references: EMBL:U86619; NID:g1840141; PID:g1840142; PIDN:AAB63523.1  
 C:Genetics:  
 A:Gene: uvsF

Query Match 6.1%; Score 105.5; DB 2; Length 1092;  
 Best Local Similarity 20.2%; Pred. No. 0.62;  
 Matches 69; Conservative 49; Mismatches 144; Indels 79; Gaps 13;

QY 26 GDKT----ATRYQEGMGAQPOENFHYRPTWDPDYEIFDPSRAIRMANWALKDPROFY 81  
 Db 705 GDQTGDYSLDYDQGMKAWEKHIILKPWDIVGKILNP--QMFSQSSTATLNDKSELY 762  
 QY 82 -----YASWATTRARQDAMESNEFEVESRRMIGLMDIDVAAR-----ALD 122  
 Db 763 FNDHEFSYLMLOENYLRTPKTRANNYEGEKRLKLE-----LADNAASSISDGLVD 815  
 QY 123 VLVLPLRHAAGANNMNA-----QICALGYGVFTAPAMFHAMDNGLGVAQYILRLALAMAE 177  
 Db 816 RMHGSQQQWLSMPTHAFISFVRPASFMFGNMMPARTSWLGQNSKOGLIIVGVIARE 875  
 QY 178 PDVLEAKATWTRDAAQPLRRYVEDTLVADPVELFIAQNALDG-----LILYPLIVY 230  
 Db 876 ADVVHAGK-----LSRYVK-----EIQGRMLRASGDRHVRQQYIPLIW 915  
 QY 231 DRFYDERTALGEGSAVAMLTAFMPEWHITESNRWIDAVVKTMAESDDNRALL-----ARW 285  
 Db 916 DRLV-KRLMWDGKSDVEDVDFMDSYFTRDDW-DALVELGLGPMDDSKVKLETQTKAAF 973  
 QY 286 TRDWSAR-----AEALAPVAARALQDAGRAALDEVRE 318  
 Db 974 TRVYNQRSHPLPYMKASNVLAAPKAPKEKPDIEDAIDESDE 1014

## RESULT 6

replication factor C protein - Emericella nidulans

Tl8306  
 replication factor C protein - Emericella nidulans  
 C:Species: Emericella nidulans, Aspergillus nidulans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: Tl8306  
 R:Kafer, E.; May, G.  
 Gene 191, 155-159, 1997  
 A:Title: The uvsF gene region in Aspergillus nidulans codes for a protein with homolo  
 A:Reference number: Z18873  
 A:Accession: Tl8306  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1092 <KAF>  
 A:Cross-references: EMBL:U86620; NID:g2258282; PID:g2258283; PIDN:AAB63574.1  
 C:Genetics:  
 A:Introns: 29/3; 38/1; 68/2  
 A>Note: uvsF

Query Match 6.1%; Score 105.5; DB 2; Length 1092;  
 Best Local Similarity 20.2%; Pred. No. 0.62;  
 Matches 69; Conservative 49; Mismatches 144; Indels 79; Gaps 13;

QY 26 GDKT----ATRYQEGMGAQPOENFHYRPTWDPDYEIFDPSRAIRMANWALKDPROFY 81  
 Db 705 GDQTGDYSLDYDQGMKAWEKHIILKPWDIVGKILNP--QMFSQSSTATLNDKSELY 762  
 QY 82 -----YASWATTRARQDAMESNEFEVESRRMIGLMDIDVAAR-----ALD 122  
 Db 763 FNDHEFSYLMLOENYLRTPKTRANNYEGEKRLKLE-----LADNAASSISDGLVD 815  
 QY 123 VLVLPLRHAAGANNMNA-----QICALGYGVFTAPAMFHAMDNGLGVAQYILRLALAMAE 177  
 Db 816 RMHGSQQQWLSMPTHAFISFVRPASFMFGNMMPARTSWLGQNSKOGLIIVGVIARE 875  
 QY 178 PDVLEAKATWTRDAAQPLRRYVEDTLVADPVELFIAQNALDG-----LILYPLIVY 230  
 Db 876 ADVVHAGK-----LSRYVK-----EIQGRMLRASGDRHVRQQYIPLIW 915  
 QY 231 DRFYDERTALGEGSAVAMLTAFMPEWHITESNRWIDAVVKTMAESDDNRALL-----ARW 285  
 Db 916 DRLV-KRLMWDGKSDVEDVDFMDSYFTRDDW-DALVELGLGPMDDSKVKLETQTKAAF 973  
 QY 286 TRDWSAR-----AEALAPVAARALQDAGRAALDEVRE 318  
 Db 974 TRVYNQRSHPLPYMKASNVLAAPKAPKEKPDIEDAIDESDE 1014

## RESULT 7

S3962  
 endoglucanase - Erwinia carotovora  
 C:Species: Erwinia carotovora  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 08-Oct-1999  
 C:Accession: S39962  
 R:Cooper, V.J.C.; Salmond, G.P.C.  
 Mol. Gen. Genet. 241, 341-350, 1993  
 A:Title: Molecular analysis of the major cellulase (CelV) of Erwinia carotovora: evid  
 A:Reference number: S39962; MUID:94067016  
 A:Accession: S39962  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-505 <COO>  
 A:Cross-references: EMBL:X76000; NID:g434941; PIDN:CAA53592.1; PID:g434942

Query Match 6.1%; Score 105; DB 2; Length 505;  
 Best Local Similarity 23.2%; Pred. No. 0.25;  
 Matches 68; Conservative 38; Mismatches 73; Indels 114; Gaps 17;

QY 112 MRDDVAARALDVLPLRHAAGANNMNAQI--CALGYGVFTAPAMFH-----AMNLL 162  
 Db 81 LRDD-----WGIVFRVAMTYAADGY---ISNPSLANKVEAAVAQAQSL 121

QY	163	GVAQYLTRLALAAEPDVLAAKATWTRDA-----W-----Qp	196
		: : : : :     : : : : :     : : : : :     : : : : :	
Db	122	GVYIIDHWILSDNDPNIIYKAQAKTFFAEMAGIYGGSPNVIETIANEPNGCVTWNGQIRP	181
		: : : : :     : : : : :     : : : : :     : : : : :	
QY	197	LRRYVEDTLVADPEVELFI-----AQNL--ALDGLL-----YPLVYDFREVDER	237
		: : : : :     : : : : :     : : : : :     : : : : :	
Db	182	YALEVTDTIIRSKPDNLIIVGTQSDIHDAAQNLDPDNTWYLUHFTYAGTHGQFLRDR	241
		: : : : :     : : : : :     : : : : :     : : : : :	
QY	238	I--ALEGGSAVAMLTAFMPQWHT-----ESNEWIDAVVKTMAAESDDNRAL--L	282
		: : : : :     : : : : :     : : : : :     : : : : :	
Db	242	IDYAAQSRGAII---FVSEWGTSDASNGGPELPESQWIDFL-----NREGVSW	287
		: : : : :     : : : : :     : : : : :     : : : : :	
QY	283	ARWTROWSARAERALAPVAARA-----LQDAGRALDEVREQFHARARLG	328
		: : : : :     : : : : :     : : : : :     : : : : :	
Db	288	VNWSLTDKPSAALAPGASGSGWTEQNLSGKF---VREQIRKA--GANLG	335
		: : : : :     : : : : :     : : : : :     : : : : :	

RESULT 8  
JL0101  
methane monooxygenase (EC 1.14.13.25) A beta chain precursor - Methylococcus capsulatus  
N:Alternate names: methane hydroxylase  
C:Species: Methylococcus capsulatus  
C:Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 15-Nov-1996  
C:Accession: JL0101  
R:Stainthorpe, A.C.; Murrell, J.C.; Salmond, G.P.C.; Dalton, H.; Lees, V.  
Arch. Microbiol. 152, 154-159, 1989  
A:Title: Molecular analysis of methane monooxygenase from Methylococcus capsulatus (Bath)  
A:Reference number: A90083; MUID:89373399  
A:Accession: JL0101  
A:Molecule type: DNA  
A:Residues: 1-387 <STA>  
C:Comment: Met-1 is removed after translation.  
C:Comment: Residues 2-41 were confirmed by direct protein sequencing.  
C:Comment: This enzyme is responsible for the initial oxygenation of methane to methanol  
atic and heterocyclic compounds.  
C:Comment: Methylococcus capsulatus has two forms of methane monooxygenase, a soluble an  
C:Comment: Protein A comprising three chains, alpha, beta and gamma, is a non-heme iron  
C:Comment: Protein B is a single subunit regulatory protein containing no prosthetic grou  
voprotein responsible for the transfer of reducing equivalents from NADH to protein A.  
C:Genetics:  
A:Gene: mmof  
C:Keywords: oxidoreductase  
F:2-387/Product: methane monooxygenase A beta chain #status predicted <MAT>

Query Match 6.0%; Score 103; DB 2; Length 387;  
Best local Similarity 20.1%; Pred. No. 0.27;  
Matches 77; Conservative 44; Mismatches 93; Indels 170; Gaps 20;

```

Qy 28 KTATRYQEGMMGAQPQ-----ENPH-YRPTWDDPYETDPSRAIRMANWY 72
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 44 KRLTEYALVYVYAQNADWITAGGLDWDGTQKPHGGRPSWGNE-----TTTELRTVDWF 96
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 73 ALKDPGRFYVASWATTARQODAMESNFPEVSRRMIGLDRDDVAARALLVPLRHAA- 131
   : | | : : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 97 KHRDPLRHPAPYVKDAE-----EWRYTDRFLQYSADGOIRAMN---PTWTS 144
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 132 ---WGAMMNAQICALGYGVFTAPAM-----FHAMDNLGVAQ--YLTR 170
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 145 CNRYWGAFLE-----YG-LFNASHQGRALSVDTVKSLAFWGFKDIDIAQMQUER 197
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 171 LALAMAPPDVLEAA---KATWTRDAAWQPLRRYVEDTLVVADPVELFIAQNLALDGL--- 224
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 198 GFLAKIVPGFDESTAVPKAETNG-----EYVKARLAVELQWE 237
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 225 -----LYPLVDRFDV---BRIALEGGSAMVLTAFM----- 253
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 238 VFDWNESAFSVHVAYDALFCQFVRREFFORLAPRFGD---NLTPFINGQYTFQIAKQ 294
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 254 -----PEWHTESNRWIDAVVTKMAAESDDNRALLARWTRDWSAREAAALPYA 301
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 295 VQDLYYNCLGDDEPF-----SDYNRTVMNRNITGKW-----LEPTI 329
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

QY 302 ARAAQD-----AGRAALDEV 316
      | | | |
      | | | |
Db 330 A-ALRDFMGLFAKLPAAGTTDKKEI 352

RESULT 9
G70763
Probable maltooligosyltrehalose trehalohydrolase - Mycobacterium tuberculosis (strain
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 29-Sep-1999
C:Accession: G70763
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987

```

Query Match 5.7%; Score 99; DB 2; Length 580;  
Best Local Similarity 19.8%; Pred. No. 0.99;  
Matches 67; Conservative 37; Mismatches 108; Indels 126; Gaps 14;

Qy	54	PDYELFDPSSAIRM-----ANWY-----	ALKDPQC 75
Db	2	PEFVWAPKPAVLRLDNGVAVHATRSADGQWHTTVAAPADARYGYLLDDDTVLPVLPDPRS 61	
Qy	80	FYASWATTRARQODAMESNFEFVESRMIGLRDDVAARALDVLPLRHAAM--GAMMNN 138	
Db	62	-----ARQPDGVHANSORPEPPQGEAARTD-----TGPEGRSVEG 97	
Qy	139	AQICALGCVTFTAPAMFHAMDNLGVAQYLTRLALAMAEPDVLEAAKATWTRDAAWQPLR 198	
Db	98	AVITELHIGTETTAGTDDAAIEKL---DYLVLDGIDEVE----- 133	
Qy	199	RYVEDTIVVADPVELEFI--AQNALDGLLYPLVYD-----RFVD-----ER 237	
Db	134	-----LMPVNSEAGTRGWDGYGLWYSVHEPYGPGDGLVRFIDACHARRLGLVLDA 184	
Qy	238	IALEGGSAVAMLTAFMPEFWHTESNRWIDAVKTKMAESDDNRALL-----ARWTRDWSA- 291	
Db	185	VFNHGFSGNVLPRFGPTLSSANPWGDG--INTAGADSDVEVHYIIDCALWRMRDFHADG 243	
Qy	292	-RAEAAALPVAARALQDAGRAALDEVREQFHARAARLG 328	
Db	244	LRLDAVALVDVTTAVH-----VLEELANATRWLSGGVG 276	

RESULT 10  
T30226  
polyketide synthase - Streptomyces hygroscopicus  
C:Species: Streptomyces hygroscopicus  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999  
C:Accession: T30226  
R:Apuricio, J.F.; Molnar, I.; Schwecke, T.; Koenig,  
Gene 169, 9-16, 1996  
A:Title: Organization of the biosynthetic gene cluster  
A:Reference number: Z20782  
A:Accession: T30226  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-8563 <APA>  
A:Cross-references: EMBL:X86780; NID:q987088; PID:q987088



C;Genetics:

A;Gene: DR0397

A;Map position: 1

C;Superfamily: formate dehydrogenase

Query Match 5.3%; Score 90.5; DB 2; Length 697;  
Best Local Similarity 20.7%; Pred. No. 6.6;  
Matches 64; Conservative 32; Mismatches 106; Indels 107; Gaps 15;

QY 60 DPSRAIRMANWYALKDPRQFYASWATRAHQDAMESNFE--FVESRR----- 107  
DB 53 DGSEKAVKLIG-----DPAHYTRGFCAKTVHPARONHPERPLPLRYNPKTPDPFP 107  
QY 108 -----MIGLMDDVAARALDVLPLR-----HAAW--GAMMNA 139  
DB 108 WERVTWDEALDDIAGRIHQLIAERGPQSILPHYAGNMGLEGSHVHALWRALGAEELE 167  
QY 140 QICA-----LGYGTVFTAPAMF--HA-----MDNLGVAQYLTRIALAMAPDVLE 182  
DB 168 TICASAGTAAMEVGYGTRLDVDPHARLIVLWGLNSLTHSLT-----PQI-T 218  
QY 183 AAKATWTRDAWQPLRRYVETLVVADPVELFIAQNLALDGLLPLVYDRFVDERIALEG 242  
DB 219 AARKNGAR-----VVCDFYRNRTA-----AFADHLKIRP 249  
QY 243 GSAVAMLTAFMPWHTESNRWIDAVVKTMAES-DDNEALLARWTRDWSARAERAAAPVA 301  
DB 250 GTDRAALGVWRELFV--NGWTDAYLTAQTEGVDDLRAA--AEWSPORTAEVTGLT 303  
QY 302 ARAQLQDAGR 310  
DB 304 ADEVROFAR 312

RESULT 14

G75442

ATP-dependent Clp proteinase, ATP-binding subunit ClpB - Deinococcus radiodurans (strain G75442)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C;Accession: G75442

R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896

A;Accession: G75442

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-875 &lt;WHI&gt;

A;Cross-references: GB:AE001956; GB:AF000513; NID:G6458776; PIDN:AAF10620.1; PID:G645877

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR1046

A;Map position: 1

C;Superfamily: endopeptidase Clp ATP-binding chain

Query Match 5.3%; Score 90.5; DB 2; Length 875;  
Best Local Similarity 23.6%; Pred. No. 8.9;  
Matches 78; Conservative 37; Mismatches 109; Indels 107; Gaps 17;

QY 76 DPROFYASWATTRARQDAMES--NFEVESKRMIGLMRDDVAARALDVLPLRHAAWGA 134  
DB 25 NPERFTASAVAINAAQLAQENHONLTHFVLRTLDNDTSAKTLA-----GG 76  
QY 135 NMN-----NAQICAL-----GYGTVFTAPAM--FHAMNLG-----VAQYLTRIAL 173  
DB 77 DLNTRAAALDAETAKLPKVQSGDGLYLDPALNRAFQKADTLAAQLGDSFVAADTLILAL 136  
QY 174 -----AMAEPOVLENAKATW-----TFDAWQWQ-- 196

DB 137 RGEYRGRLPDEVSLNRAVTEQRKGKTVINKTSEQFDALNKYGTDLTQRAEDGKFDPI 196  
QY 197 -----LRRYVEDTLVADP--VELFIAQNLAL-----DGLLYPLVYDRF 233  
DB 197 GRDEIRRMVQILLRRSKNNPNVLIGEIPGVGKTAIAEGIAMRIVKGVDPGLR----- 248  
QY 234 VDERI-ALEGGSAVAMLTAFMPWHTESNRWIDAVVKTMAAE-----SDNRALLARWTRD 288  
DB 249 -DKKIVSLEMGLLAG-AKFRGEFERUKGVIDEVVK-SAGEIILFVDEIHTIVGAGKTE 305  
QY 289 WSAARAALAPVAARA-LQDAGRAALDEVRE 318  
DB 306 GSPDAGNMLKPALARGELHLIGATTILDEVRE 336

RESULT 15

A58538

4-hydroxybenzoate--CoA ligase (EC 6.2.1.27) - Rhodopseudomonas palustris

C;Species: Rhodopseudomonas palustris

C;Date: 31-Mar-1997 #sequence\_revision 11-Apr-1997 #text\_change 18-Feb-2000

C;Accession: A58538; A36960

R;Egland, P.; Harwood, C.S.

submitted to the Protein Sequence Database, March 1997

A;Description: Rhodopseudomonas palustris 4-hydroxybenzoate coenzyme A ligase precurs

A;Reference number: A58538

A;Contents: Correction

A;Accession: A58538

A;Molecule type: DNA

A;Residues: 1-539 &lt;EGI&gt;

A;Cross-references: GB:U02033; NID:G665919; PIDN:AAA62604.1; PID:G665920

R;Gibson, J.; Dispensa, M.; Fogg, G.C.; Evans, D.T.; Harwood, C.S.

J. Bacteriol. 176, 634-641, 1994

A;Title: 4-hydroxybenzoate-coenzyme A ligase from Rhodopseudomonas palustris: purified

A;Reference number: A36960; MUID:94131941

A;Accession: A36960

A;Status: significant sequence differences

A;Molecule type: DNA

A;Cross-references: GB:U02033; NID:G665919

C;Genetics:

A;Gene: hbaA

C;Superfamily: human SA protein; acetate--CoA ligase homology

C;Keywords: acid-thiol ligase; aromatic hydrocarbon catabolism

F:54-509/Domain: acetate--CoA ligase homology &lt;ACL&gt;

Query Match 5.2%; Score 90; DB 2; Length 539;

Best Local Similarity 20.8%; Pred. No. 5.2;

Matches 65; Conservative 32; Mismatches 93; Indels 122; Gaps 16;

QY 107 RMIGLMRDDVAARALDVLPLRHAAGANMNAQICALGYGT----- 148  
DB 198 RIIGYREDDDVVSAAKLFF-----AYG--LGNMFCPMGIGATSVLYPERPTADSVFDTL 250  
QY 149 -----VFTAPAMPHAM-----DNLGVAQYLTRIALAMAEF----- 178  
DB 251 RLHQPTLLFAVPTLYAAMLADPRSRTETLPDL-----RLCVSAGEPLPAQVGLNWRN 303  
QY 179 ---DVLAAKATWTRDAWQPLRRYVEDTLVWADPVELFIAQNLALDGLLPLVYDR-- 232  
DB 304 RFGHDIVNGVSGSTEMGHFLFLNLPHAVE-----YGTSGVPVDGYRLRLYGDGRGQ 352  
QY 233 -FVDERIA---LEGGSAVA-----MLTAFMPWHTESNRW--IDAVVKTMAESDD 277  
DB 353 DVADEIGELLVSGSSAAGYWNQDKRTTFVGEWTRTGDKYHRRADGVY-TYCGRTDD 411  
QY 278 NRALLARWTRDWSAR-----AEALAP-----VAARALQDAGRAAL 313  
DB 412 LFKVSGIWSVPFEIQALMSHAKVLEAAVIPAEDTDGLIKPKAFIVLASRGDIDFG-ALF 470  
QY 314 DEVREOFHARAA 325  
DB 471 DELKE--HVKSA 480



Search completed: September 26, 2000, 20:24:12  
Job time: 3636 sec

---

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:22:56 ; Search time 58.11 Seconds  
(without alignments)  
176.517 Million cell updates/sec

Title: US-09-430-029-3

Perfect score: 1723

Sequence: 1 MTIELKTVDIKPLRTFFAHV.....ALDEVREQFHAPAAALGIAL 331

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	779.5	45.2	331	1	DMPL_PSESP
2	184	10.7	326	1	TMOE_PSEME
3	123	7.1	394	1	MEMB_METTR
4	106	6.2	444	1	GUNN_ERWCA
5	105	6.1	505	1	GUNV_ERWCA
6	103	6.0	386	1	MEMB_METCA
7	99	5.7	580	1	YF62_MYCTU
8	91.5	5.3	472	1	Y568_MYCTU
9	88.5	5.1	461	1	KLB2_ECOLI
10	88.5	5.1	504	1	GUNW_ERWCA
11	88.5	5.1	509	1	GATB_MYCLE
12	87.5	5.1	1396	1	VLTF_BPT5
13	87	5.0	573	1	Y2G8_MYCLE
14	86.5	5.0	454	1	RHSE_RHIME
15	86.5	5.0	554	1	IJUK_ECOLI
16	85.5	5.0	474	1	UCR2_FUGGR
17	85.5	5.0	509	1	GATB_MYCTU
18	85.5	5.0	1275	1	REBC_MVXXA
19	85	4.9	456	1	REP_STRLI
20	84.5	4.9	365	1	E13B_NICPL
21	84.5	4.9	370	1	E13B_TOBAC
22	84.5	4.9	371	1	E13B_TOBAC
23	83	4.8	416	1	PROA_STNTR
24	83	4.8	573	1	Y2G8_MYCTU
25	82.5	4.8	572	1	SYM_AERPE
26	82.5	4.8	1278	1	DHBE_BACSU
27	82	4.8	429	1	RSI_LEULA
28	81.5	4.7	383	1	MEXA_PSEAE
29	81.5	4.7	544	1	CH61_SYNVU
30	81.5	4.7	1234	1	MF0_MYCTU
31	81	4.7	154	1	YD88_SYNY3
32	81	4.7	451	1	Y0D0_MYCTU
33	81	4.7	559	1	ILVB_KLEPN

34	81	4.7	918	1	IMB2_YEAST	P38217	saccharomyc
35	80.5	4.7	532	1	IPA7_SHIFL	P18014	shigella fl
36	80	4.6	541	1	CH60_CYAPA	Q37757	cyanophora
37	80	4.6	810	1	CLPC_BACSU	P37571	bacillus su
38	79.5	4.6	339	1	RFAT_ECOLI	P27128	escherichia
39	79.5	4.6	574	1	IPAA_SHIFL	P18009	shigella fl
40	79.5	4.6	718	1	CTPC_MYCTU	P96875	mycobacteri
41	79.5	4.6	1829	1	FRPC_NEIME	P5127	neisseria m
42	79	4.6	547	1	SAPA_ECOLI	Q47622	escherichia
43	79	4.6	657	1	CTPD_MYCTU	O53160	mycobacteri
44	79	4.6	759	1	YEHM_ECOLI	P33349	escherichia
45	79	4.6	947	1	SECA_STRLI	P55021	streptomyce

## ALIGNMENTS

RESULT 1  
DMPL\_PSESP

ID	DMPL_PSESP	STANDARD;	PRT;	331 AA.
AC	P19730;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	PHENOL HYDROXYLASE P1 PROTEIN (EC 1.14.13.7) (PHENOL 2-MONOOXYGENASE			
DE	P1 COMPONENT).			
GN	DMPL OR PHEA2.			
OS	Pseudomonas sp. (strain CF600).			
OG	Plasmid pVII50.			
OC	Bacteria; Proteobacteria.			
RC	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 91072230.			
RA	Nordlund I., Powlowski J., Shingler V.;			
RT	*Complete nucleotide sequence and polypeptide analysis of			
RT	multicomponent phenol hydroxylase from Pseudomonas sp. strain			
RT	CF600.*;			
RL	J. Bacteriol. 172:6826-6833(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BH;			
RA	Takeo M., Maeda Y., Okada H., Miyama K., Mori K., Ike M.,			
RA	Fujita M.;			
CC	Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED			
CC	DERIVATIVES. P1 IS REQUIRED FOR GROWTH ON PHENOL, AND FOR			
CC	IN VITRO PHENOL HYDROXYLASE ACTIVITY.			
CC	-!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) -> CATECHOL + NADP(+)			
CC	+ H(2)O.			
CC	-!- COFACTOR: FAD, AND REQUIRES FE(+2) FOR ACTIVITY.			
CC	-!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.			
CC	-!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED			
CC	BY P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.			

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
EMBL: M60276; AAA25940.1; -  
EMBL: D28864; BAA06015.1; -  
PIR: B37831; B37831.  
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;  
NADP; Flavoprotein; FAD; Iron; Plasmid.  
SQ SEQUENCE 331 AA; 38208 MW; 8042F5723BE3A5E8 CRC64;

Query Match 45.2%; Score 779.5; DB 1; Length 331;  
Best local similarity 46.8%; Pred. No. 5.3e-60;  
Matches 155; Conservative 54; Mismatches 115; Indels 7; Gaps 4;



[illegible]



RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO  
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.  
 CC -!- SIMILARITY: TO M.TUBERCULOSIS RV1563C.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Z74020; CAA98329.1; -  
 DR TUBERCULIST: RV1562C; -  
 DR PRAM: PF00128; alpha-amylase; 1.  
 KW Hypothetical protein; Hydrolase; Glycosidase.  
 SQ SEQUENCE 580 AA; 64076 MW; 506190468F4B862 CRC64;

Query Match 5.7%; Score 99; DB 1; Length 580;  
 Best Local Similarity 19.8%; Pred. No. 0.43; 108; Indels 126; Gaps 14;  
 Matches 67; Conservative 37; Mismatches 108; Indels 126; Gaps 14;  
 QY 54 PDYEIFDSRAIR- - - - - ANWY- - - - - ALKDPK 79  
 D 2 PEFRWAFKPAVLVDVNGAVHAMTRSDGWHHTVAADARYGLLDDPTVLPDPRS 61  
 QY 80 FYASWATTRAQDAMESNEFEVSRMIGLRRDDVAARALDVLPLRHAW-GANNNN 138  
 D 62 - - - - - ARQPDGVHARSQRPEPPQFGAARTD - - - - - TCWPGRSVEG 97  
 QY 139 AQICALGYGTFTAPAMFAMDNLGVAOYLRLALAMAEPPDVLAEAKATWTDRAWQPIR 198  
 D 98 AVIVELHIGTFTAGTTPDAIEKL- - - - - DYLVDLGIDFVE- - - - - 133  
 QY 199 RYVEDTLVVADEVLELFT-AQNIALDGLLYPLVYD- - - - - RFVD- - - - - ER 237  
 D 134 - - - - - LMPVNSFAGTRGNGYDGLWYSVHEFYGGPDGLVRFIDACHARRLGLVIDA 184  
 QY 238 IALEGGSVAMLTAFMPWHHTESNRWIDAVVKTMAESDDNFALL- - - - - ARWTRWSA- - - 291  
 D 185 VFNHLGPSNGYLPREFGPYISSASNPWSDG-INIAGADSDEVRHYIIDCALRWMRDFHADG 243  
 QY 292 -RAEALAPVAARALQDAGRALDVRQEFHARAARLG 328  
 D 244 LRLDAVHALVDTTAVH- - - - - VLELANATRWLSGQLG 276

RESULT 8  
 Y568\_MYCTU  
 ID Y568\_MYCTU STANDARD; PRT; 472 AA.  
 AC OS3765;  
 DT 15-FEB-2000 (Rel. 39, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last sequence update)  
 DE PUTATIVE CYTOCHROME P450 RV0568 (EC 1.14.-.-).  
 GN RV0568 OR MT039.06.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-H37RV;  
 RX MEDLINE; 98295987.  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,  
 RA Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AL021942; CAA17439.1; -  
 DR TUBERCULIST: RV0568; -  
 DR PRAM: PF00067; P450; 2.  
 DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
 KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme.  
 FT BINDING 388 388 HEME (BY SIMILARITY).  
 SQ SEQUENCE 472 AA; 50687 MW; B7F2898BEE80863F CRC64;

Query Match 5.3%; Score 91.5; DB 1; Length 472;  
 Best Local Similarity 25.2%; Pred. No. 1.5;  
 Matches 41; Conservative 20; Mismatches 69; Indels 33; Gaps 4;  
 QY 162 LGVAOYLRLALAMAEPPDVLAEAKATWTDRAWQPIRKYVEDTLVVADEVLELFT 221  
 D 177 LNVGSPWAT- - - LALANPSLL- - - - - NNRLWSLRRIIEEA- - - - - 208  
 QY 222 DGLLYPLVYDREVDIERIALEGGSVAMLTAFMPWHHTESNRWIDAVVKTMAESDDNRA- 280  
 D 209 DALLVLAETADRADPDLAARTDTLAMLVRAADEGDTWTETRELROOLITLVAGHTTAT 268  
 QY 281 - - - - - LLARWTRWSARAALAPVAARALQDAGRALDVRQEF 319  
 D 269 GLSWALERLTRHPVTLAKAVQAADASAAGDPAGDEYIDAVAKE 311

RESULT 9  
 KLB2\_ECOLI  
 ID KLB2\_ECOLI STANDARD; PRT; 461 AA.  
 AC P52605;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE KLCB PROTEIN.  
 GN KLCB.  
 OS Escherichia coli.  
 OG Plasmid Incp-alpha RK2.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 94285211.  
 RA Pansegrau W., Lanka E., Barth P.T., Figurski D.H., Guiney D.G.,  
 RA Haas D., Helinski D.R., Schwab H., Stanisich V.A., Thomas C.M.,  
 RT "Complete nucleotide sequence of Birmingham Incp alpha plasmids.  
 Compilation and comparative analysis.";  
 RL J. Mol. Biol. 239:623-663(1994).  
 RN [2]  
 RP SEQUENCE OF 1-237 FROM N.A.  
 RA Larsen M.H., Figurski D.H.;  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 CC -!- DOMAIN: CONTAINS A DNA-BINDING REGION JOINED BY A SHORT VARIABLE  
 CC SEGMENT TO A REGION SIMILAR TO E-COLI KORA AND TRBA.

CC -----  
 CC this SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L27758; -- NOT\_ANNOTATED\_CDS.  
 DR EMBL; U05773; AA57449.1; --  
 KW Plasmid; DNA-binding; Transcription regulation.  
 FT SIMILAR 315 344 TO THE C-TERMINAL OF E.COLI KORA AND  
 FT TRBA.  
 SQ SEQUENCE 461 AA; 51165 MW; 10EBFB2F2E00BD81 CRC64;

Query Match 5.1%; Score 88.5; DB 1; Length 461;  
 Best Local Similarity 22.5%; Pred. No. 2.6;  
 Matches 56; Conservative 31; Mismatches 73; Indels 89; Gaps 14;

QY 90 ARQDAMESNEFEVESRMIGMRDDVAARALDVLPLRHAAG-----AMNNAQI-- 141  
 DB 171 AGQDAAE-----RVI-----ERCRATPGVPM-----WGQEGDFLASVDGMKRVWV 212  
 QY 142 -CALGYGTVTAPAMFAMD-----NLGVAQYLTRLALAMAEPDVLEAAKATWTRDAA 193  
 DB 213 EVESYGGTLTVHPQFHAVLDLDPFISETGYRSHYDHARGMTVDQVADGVLRALLSH- 271  
 QY 194 WQPLRVVDETLVADPVELFIAQNLADGLLYPLVDFRVDERIALEGGSVAVAMLT--- 250  
 DB 272 -----RRYID-----ARD-----QDLADEPLP-----AWLAGITPPP 299  
 QY 251 -----AFMPEWHTESN-----RWIDAVVKTMAESDDNRALLARWTRDWSARAALAPVA 301  
 DB 300 RRVAVVEDMKPDELPGFAWVDVLP-----HQAFIA-----RKWAASAKAKLAAR 349  
 QY 302 ARALQDAGR 310  
 DB 350 AKAQEPAGQ 358

RESULT 10  
 GUNW\_ERWCA STANDARD; PRT; 504 AA.  
 AC Q59395;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE ENDOGLUCANASE V1 PRECURSOR (PC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE V1)  
 DE (CELLULOSE V1).  
 GN CELV1.  
 OS Erwinia carotovora.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Erwinia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SCC3193;  
 RX MEDLINE: 95231512.  
 RA Mae A., Heikinheimo R., Palva E.T.;  
 RT "Structure and regulation of the Erwinia carotovora subspecies  
 RT carotovora SCC3193 cellulase gene celv1 and the role of cellulase in  
 RT phytopathogenicity.";  
 RL Mol. Gen. Genet. 247:17-26(1995).  
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC  
 CC LINKAGES IN CELLULOSE.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL  
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X79241; CAA55823.1; --  
 DR HSSP; Q06851; LNBC.  
 DR PFAM; PF00942; CBD\_3; 1.  
 DR PFAM; PF00150; cellulase; 1.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
 FT SIGNAL 1 31 POTENTIAL.  
 FT CHAIN 32 504 ENDOGLUCANASE V1.  
 FT DOMAIN 32 334 CATALYTIC.  
 FT DOMAIN 335 352 LINKER.  
 FT DOMAIN 353 504 CELLULOSE-BINDING (BY SIMILARITY).  
 FT ACT\_SITE 168 188 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 256 266 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 504 AA; 54963 MW; 0D7ECF74781565FA CRC64;

Query Match 5.1%; Score 88.5; DB 1; Length 504;  
 Best Local Similarity 23.4%; Pred. No. 2.9;  
 Matches 65; Conservative 31; Mismatches 99; Indels 83; Gaps 16;

QY 112 MRDDVAARALDVLPLRHAAGCAN-----MNAQICALGYCTVETAPAMFAMDNLGVA 165  
 DB 81 LRDDGIVNFRVAMTYAENGTYIANPSLANKYKVAAGGLGVIID--MHTLSDNPN 138  
 QY 166 QYLTRLALAMAE-----PDVL-EAA-----KATWTRDAWQPLRRYVEDTLVADPV 211  
 DB 139 TYKAQAKITFEEMAGLYGNSPNVIYEIANEPNGSVTW--NGQIRPYALEVDTITRSKQPD 196  
 QY 212 ELFI-----AQNALD--GLLYPL-----VYDFVDERI--ALEGGSVAVAMLT 250  
 DB 197 NLIIVGSGTWSQDTHDAADNQLPDPNTLYALHFYAGTHQGLRDRIDYQAQSGAAI---- 252  
 QY 251 AFMPEWHT-----ESNRWIDAVVKTMAESDDNRAL-LARWTRDWSARAAL 297  
 DB 253 -FVSEWGTSDASNGGPPFPESQWIDFL-----NNRGISWNWSISDKSSTAAL 302  
 QY 298 APVAARA-----IQDAGRAALDEVREQFHARAALRG 328  
 DB 303 VAGASKSGGWTEQNLSTSGKE----VREQIRAGAGLSG 336

RESULT 11  
 GATB\_MYCLE STANDARD; PRT; 509 AA.  
 AC O33107;  
 DT 15-FEB-2000 (Rel. 39, Created)  
 DT 15-FEB-2000 (Rel. 39, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT B (EC 6.3.5.-) (GLU-ADT  
 DE SUBUNIT B).  
 GN GATB OR MLCB637.15.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Oliver K., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases  
 CC -1- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED  
 CC GLN-TRNA(GLN) THROUGH THE TRANSAMIDATION OF MISACYLATED GLU-  
 CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE  
 CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH  
 CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP  
 CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.  
 CC -1- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE GATB FAMILY.  
 CC -----





Query Match 5.0%; Score 87; DB 1; Length 573;  
 Best Local Similarity 22.8%; Pred. No. 4.5;  
 Matches 82; Conservative 37; Mismatches 123; Indels 118; Gaps 17;

QY 25 GCDKATRYQEGMGAQPOENF-----HYRPT-----WDPYEIFDPSRS 64  
 DB 220 GNEDAAVALLEWLTTHPEKVSAAKDPVYRLTTTAAQIAARADPDPSPSWTDSNR 279  
 QY 65 AIRMANWVALKDPPOFYFASWATTRARQODAMESFEVESRMIGLRRDDVAARALDVL 124  
 DB 280 DRLTQAQALD-RQI-----GLTRVKIQ-----IERYAATMMAKVAAGMKVA 324  
 QY 125 VPLRHAANGANNNAQICALGYGTFTAPAMPAMDNLGVAQYTLRLALAMAPDVLAAA 184  
 DB 325 OPSKHMIEGT-----PPGFGKTTIARVVANILAGLV-----ISEPKLIVETS 366  
 QY 185 KATWTRDAWQPLRYVEDTLVADPVELFIAQNALDGLL-----YPLVDRFVDERIA 239  
 DB 367 RQDFVAETEGQ-----SAVRAAKTIDIALGVLFDIDAYALVOER--DGRTD 411  
 QY 240 LEGGSAAVLAFTAFPEWHTESNRWIDAVVKTMAESDD-----NRALLARW-TR--- 287  
 DB 412 PFGQEAULTLARM-----ENDR--DRLVVIAGYSSDIDRLLETNEGLRSFATRIEFD 464  
 QY 288 -----DWSARAE-----ALAPVAARALODAGRAALDEVREQFHAR 323  
 DB 465 TYPDELEIAKVIATDADSLSAEASKNLEAAKQLAQLTLR--GRPALDVAGNGRYAR 522

RESULT 14  
 RHSE\_RHIME STANDARD; PRT; 454 AA.  
 AC Q923Q8;  
 DT 15-FEB-2000 (Rel. 39, Created)  
 DT 15-FEB-2000 (Rel. 39, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE RHIZOACTIN SIDEROPHORE BIOSYNTHESIS PROTEIN RHSE (EC 1.-.-.-).  
 GN RHSE.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OG Plasmid pSymA (megaplasmid 1).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RCR2011 / SU47;  
 RA Lynch D., O'Connell M., O'Brien J.;  
 RT "Cloning and sequence analysis of the Rhizobium meliloti 2011  
 rhizobactin regulon";  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- COFACTOR: FAD (BY SIMILARITY).  
 CC -1- PATHWAY: RHIZOACTIN SIDEROPHORE BIOSYNTHESIS.  
 CC -1- SIMILARITY: BELONGS TO THE LYSINE N6-HYDROXYLASE / L-ORNITHINE N5-  
 OXYGENASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF110737; AD09416.1; -;  
 KW Oxidoreductase; Iron transport; NADP; Flavoprotein; FAD; Plasmid.  
 FT NP\_BIND 7 13 FAD (POTENTIAL).  
 SQ SEQUENCE 454 AA; 50211 MW; 57074278FFAFC1D1 CRC64;

Query Match 5.0%; Score 86.5; DB 1; Length 454;  
 Best Local Similarity 23.0%; Pred. No. 3.7;  
 Matches 66; Conservative 40; Mismatches 102; Indels 79; Gaps 16;

QY 28 KTATRYQEGM--GAQPOENF--HYRPTWDPDYEIFDPSRSAL--RMANWYALKD---PR 78  
 DB 36 KPAFRHHEGLIIPGTLQVFPFADLVMTADPTRLUSFLNVLAVHDLRLYKFFYFENPMIR 95  
 QY 79 QFY--YASWATTRARQODAMESFEVESRMIGLRRDDVAARAL--DVLVPLRHAANGA 134  
 DB 96 QEYDHYCRWA--SQLSACRFGEEV-----DVAHESASDSFIVESRSASGSK 141  
 QY 135 NMNNAQICALGYGT-----VFTAPAMFAMD-----NLGVAQYLTR 170  
 DB 142 QOYRSNTAIGVGTAFPLPKWAIKITLAPIMHSEFGRRLRLSKRRRVTVIGSGSAAE 201  
 QY 171 LALANAE--PDVLEA-AKATW--TRDAWQ-----LRRYVEDTL----- 205  
 DB 202 CVIALNLDTPEWVAAGASIQWITRSAGFFPMEYSKLGLEYTTPDMRHFHRIAPVRRRE 261  
 QY 206 VVADPVELFIAQNALDGLLYPLVDRFY---DERIALEGSSAVML 249  
 DB 262 IVADQGLLYKGISFTSTIGTIFDLMYERSVGGDRDPGLALFSCAVETL 308

RESULT 15  
 YQJK\_ECOLI STANDARD; PRT; 554 AA.  
 ID YQJK\_ECOLI  
 AC P37797;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-FEB-2000 (Rel. 39, Last annotation update)  
 DE ABC TRANSPORTER ATP-BINDING PROTEIN YQJK.  
 GN YQJK.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE; 95334362.  
 RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,  
 RA Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome VI: DNA sequence of the  
 RT region from 92.8 through 100 minutes.";  
 RL Nucleic Acids Res. 23:2105-2119(1995).  
 RN [2]  
 RP REVISIONS TO C-TERMINUS.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE; 97426617.  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE OF 1-79 FROM N.A.  
 RX MEDLINE; 92041559.  
 RA Engel H., Kazemier B., Keck W.;  
 RT "Murein-metabolizing enzymes from Escherichia coli: sequence analysis  
 RT and controlled overexpression of the slt gene, which encodes the  
 RT soluble lytic transglycosylase.";  
 RL J. Bacteriol. 173:6773-6782(1991).  
 RN [4]  
 RP PARTIAL SEQUENCE OF 1-11.  
 RC STRAIN=K12 / EMG2;  
 RX MEDLINE; 97443975.  
 RA Link A.J., Robison K., Church G.M.;  
 RT "Comparing the predicted and observed properties of proteins encoded  
 RT in the genome of Escherichia coli K-12.";  
 RL Electrophoresis 18:1259-1313(1997).  
 RN [5]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RX MEDLINE; 99420866.  
 RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;  
 RT "Enrichment of low abundance proteins of Escherichia coli by

```
RT hydroxyapatite chromatography." ;
RL Electrophoresis 20:2181-2195(1999).
CC -|- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC (ABC TRANSPORTERS). EF-3 SUBFAMILY. STRONG, TO H.INFLUENZAE
CC HI1252.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14003; AAA97287.1; ALT_FRAME.
DR EMBL; AE000509; AAC77344.1; -.
DR EMBL; M69185; -; NOT_ANNOTATED_CDS..
DR ECOGENE; EG12343; YUJK.
DR FFAM; PF00005; ABC_tran; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW ATP-binding; Transport; Repeat.
FT INIT_MET 0
FT NP_BIND 38 45 ATP (POTENTIAL).
FT NP_BIND 355 362 ATP (POTENTIAL).
SQ SEQUENCE 554 AA; 62311 MW; 444C605C67A019B1 CRC64;
```

Query Match 5.0%; Score 86.5; DB 1; Length 554;  
Best Local Similarity 17.8%; Pred No. 4.8; Mismatches 129; Gaps 14;  
Matches 67; Conservative 50;

QY	10	IKPLRHTFAHVA-----ONIGGDKTATRYQEGM-----MGAQ 41
Db	14	VPPKRHIILKNISLSPFGAKIGVLGLNGAGKSTLLRIMAGIDKIDIEGEARPOPDIKIGYL 73
QY	42	PQENFHYRTWPDYELFPDSAI-----RMANWTALKDPROFYASWATTTARQQ 93
Db	74	POE-----PQLNPEHTVRESIEEAVSEVVNALKRDLDEVYALYADPDADFKLAAEQGRLE 128
QY	94	DAMESNFEEVESRRMIGLRDVAARALDVLVPLRHAAGANNNAQICAIQGYGTVFTAP 153
Db	129	ELIQAH-----DGHNLNVQLERAADA---LRLPDWDAKIANL----- 162
QY	154	AMFHAMDNLGVAQYLTRLALAMAEVDVLEAAKATWTRDA---AWQPLRRYVEDTLVYADP 210
Db	163	-----SGGERRRVALCRILLEKPDMLLDDEPTNHLDAESVAV--LERFLHD----- 206
QY	211	VELFTAQNLALDGLLYPLVYDRFVDERIA-----LEGGSVAVMLTAFMPFWHTESNRWID 265
Db	207	-----FEGTVVAITHDRYFLDNVAGWILELDRGEGI-----PWEGNYSWLE 248
QY	266	AVVKTMAESDDNRLALRWTRDWSARAFAALAPVAARALQDAGRAAL----- 313
Db	249	OKDQRLAQASQEA-----RRKSIEKELEWVRQGTGKQSGKARLARFEELNSTEYQ 302
QY	314	--DEVREQFHARAARG 328
Db	303	KRNETNELEFIPGPRLG 319

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:11:41 ; Search time 75.19 Seconds  
(without alignments)  
28.036 Million cell updates/sec

Title: US-09-430-029-4

Perfect score: 451  
Sequence: 1 MSNVFIAFOANESRPIVDA.....ITLSGHIDEDEFTLSWSH 89

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	355	78.7	92	1 W06801	Toluene ortho-mono
2	257	57.0	90	1 W98972	Alcaligenes sp. pr
3	67.5	15.0	410	1 R51368	Protein containing
4	65	14.4	1093	1 R40815	Truncated IBV spik
5	65	14.4	1162	1 P70137	Sequence of a regi
6	65	14.4	1162	1 W06829	Major neutralising
7	64.5	14.3	394	1 W97744	Wheat S-adenosylme
8	64	14.2	292	1 W80988	Amino acid sequenc
9	64	14.2	1162	1 P60720	Sequence of the SI
10	63.5	14.1	394	1 W34540	S-adenosylmethioni
11	63.5	14.1	394	1 W34541	S-adenosylmethioni
12	63.5	14.1	394	1 W34542	S-adenosylmethioni
13	62.5	13.9	406	1 R11349	Cytochrome enzyme
14	62	13.7	322	1 R61118	Yeast presequence
15	62	13.7	897	1 R20982	Sequence of beta-c
16	61	13.5	256	1 W01203	Serine protease pf
17	60.5	13.4	333	1 R39490	Human apoAIV mutel
18	60.5	13.4	342	1 R39491	Human apoAIV mutel
19	60.5	13.4	1050	1 W67642	A serine/threonine
20	60.5	13.4	1168	1 R11611	Splice protein of 1
21	60.5	13.4	2639	1 R73476	Grapevine leafroll
22	59.5	13.2	339	1 R89748	AFT-1 interacting
23	59.5	13.2	464	1 W25116	CD2-associated int
24	59.5	13.2	464	1 W26496	CD2 associated int
25	59.5	13.2	464	1 W80420	CD2 associated int
26	59.5	13.2	553	1 W25115	CD2-associated int
27	59.5	13.2	553	1 W26495	CD2 associated int
28	59.5	13.2	553	1 W80419	CD2 associated int
29	59.5	13.2	659	1 W37724	CD2 associated int
30	58.5	13.0	732	1 R11056	Acylamino acid-iso
31	58	12.9	117	1 R66214	Protein encoded by
32	58	12.9	117	1 R81470	Nocardia corallina
33	58	12.9	681	1 R72970	GLVR-1 protein. Ne

34 58 12.9 681 1 W96999 The gibbon ape leu  
35 58 12.9 1477 1 R67691 S. cerevisiae scau  
36 58 12.9 1477 1 W10424 Saccharomyces cere  
37 58 12.9 1477 1 W06819 Fumonisin-resistan  
38 57.5 12.7 504 1 W36140 Bovine p58 protein  
39 57 12.6 1051 1 W67641 A serine/threonine  
40 56.5 12.5 297 1 Y07078 Renal cancer assoc  
41 56.5 12.5 355 1 W29999 Polyamide hydrolas  
42 56.5 12.5 355 1 W36011 Polyamide hydrolas  
43 56.5 12.5 618 1 R94587 Dnak protein. Dnak  
44 56 12.4 459 1 Y11069 H. pylori ORF 12ae  
45 56 12.4 559 1 W19630 Streptomyces venez

## ALIGNMENTS

## RESULT 1

W06801  
ID W06801 standard; Protein; 92 AA.

AC W06801;

DT 29-JAN-1997 (first entry)

DE Toluene ortho-monoxygenase subunit tomA2.

KW tom; prom; self-transmissible; constitutive; bioreactor; pollutant;

KW breakdown; trichloroethylene; TCE; degradation.

OS Pseudomonas cepacia strain PRL-23.

PN US5543317-A.

PD 06-AUG-1996.

PF 02-MAY-1991; 694718.

PR 02-MAY-1991; US-694718.

PR 15-DEC-1993; US-167457.

PR 06-OCT-1994; US-319387.

PA (FRAN/) FRANCESCONI S C.

PA (SHIE/) SHIELDS M S.

PI Francesconi SC; Shields MS;

DR WPI; 96-370640/37.

DR N-PSDB; T44457.

PT Microorganisms transformed with P. cepacia PRL-23 Tom enzyme gene -  
Claim 1: Column 29-30; 25pp; English.

PS The present sequence is that of toluene ortho-monoxygenase subunit

CC tomA2 encoded by T44457, isolated from Pseudomonas cepacia strain

CC PRL-23. The Tom gene is present on a large self-transmissible plasmid

CC denoted PTOM. The enzyme is capable of degrading trichloroethylene

CC (TCE), a hazardous pollutant. The PTOM plasmid is transmissible and

CC expressible in other bacteria, thus many bacteria can be genetically

CC altered to constitutively degrade TCE, esp. in bioreactors or

CC TCE-contaminated environments. P. cepacia PRL-23 contg. PTOM does not

CC need exogenous chemical inducers and is capable of functioning under a

CC diverse set of conditions. Also it does not require an inducer that is a

CC co-substrate for the enzyme required to break down TCE, in effect, it is

CC not subject to competitive inhibition.

SQ Sequence 92 AA;

Query Match 78.7%; Score 355; DB 1; Length 92;

Best Local Similarity 73.0%; Pred. No. 1.9e-38;

Matches 65; Conservative 14; Mismatches 10; Indels 0; Gaps 0;

QY 1 MSNVFIAFOANESRPIVDAIVDNRAVVPSPGMVKIDAPDLATRTTIEELTGRF 60

DB 3 MSNVFIAFOANESRPPVEAILADNPKAVATESPGMVKIDAPGHLTINRQSIDRIGMKF 62

QY 61 DLQQLQVNLITLSCHIDEDEFTLSWSH 89

DB 63 DLQQLHNLITLSGYIDEDDEQFTLSWKH 91

## RESULT 2

W98972

ID W98972 standard; Protein; 90 AA.

AC W98972;

DT 10-MAY-1999 (first entry)









```
PS Claim 2; Page 6-7; 13pp; Japanese.
CC This sequence represents the S-adenosylmethionine synthase 2 (sam2)
CC protein. The DNA encoding this sequence may be used in producing plants
CC which are resistant to alkaline soil.
SQ Sequence 394 AA;

Query Match 14.1%; Score 63.5; DB 1; Length 394;
Best Local Similarity 30.2%; Pred. No. 3.4;
Matches 32; Conservative 9; Mismatches 34; Indels 31; Gaps 6;

QY 2 SNVFIARQANED-----SRPIVD---AIVADNPRAVWVSPGMVKIDAPDRLTIRRET 51
Db 283 SGATIAQQAASIIASGLARLCIQAISYAGCPPEPLSVFVDSYGTGKI--PDR-----E 334
QY 52 IEELTGRFDLQ--QLOVNL-----ITLSGHIDEDEDEFT 84
Db 335 ILKLVKENFDRPGMIINLDLKKGNRFKTAAYGHGRDDAFT 380

RESULT 12
W34542
ID W34542 standard; protein; 394 AA.
AC W34542;
DE 26-MAR-1998 (first entry)
DE S-adenosylmethionine synthase 3.
KW S-adenosylmethionine synthase 3; sam3; barley; alkali resistant plant.
OS Hordeum vulgare.
PN J09313186-A.
PD 09-DEC-1997.
PF 28-MAY-1996; 133406.
PR 28-MAY-1996; JP-133406.
PA (NIOC ) NIPPON OIL CO LTD.
DR WPI; 98-080077/08.
DR N-PSDB; T99143.
PT S-adenosyl-methionine synthase gene - useful in producing plants
PT resistant to alkaline soil
PS Claim 3; Page 7-8; 13pp; Japanese.
CC This sequence represents the S-adenosylmethionine synthase 3 (sam3)
CC protein. The DNA encoding this sequence may be used in producing plants
CC which are resistant to alkaline soil.
SQ Sequence 394 AA;

Query Match 14.1%; Score 63.5; DB 1; Length 394;
Best Local Similarity 30.2%; Pred. No. 3.4;
Matches 32; Conservative 9; Mismatches 34; Indels 31; Gaps 6;

QY 2 SNVFIARQANED-----SRPIVD---AIVADNPRAVWVSPGMVKIDAPDRLTIRRET 51
Db 283 SGATIAQQAASIIASGLARLCIQAISYAGCPPEPLSVFVDSYGTGKI--PDR-----E 334
QY 52 IEELTGRFDLQ--QLOVNL-----ITLSGHIDEDEDEFT 84
Db 335 ILKLVKENFDRPGMIINLDLKKGNRFKTAAYGHGRDDAFT 380

RESULT 13
R11349
ID R11349 standard; Protein; 406 AA.
AC R11349;
DE 05-JUN-1991 (first entry)
DE Cytochrome enzyme P450SU1.
KW Cytochrome P450; P450SU1; P450SU2; herbicide resistance.
OS Streptomyces griseolus.
PN W09103561-A.
PD 21-MAR-1991.
PF 27-AUG-1990; U04785.
PR 11-SEP-1989; US-405605.
PR 12-JAN-1990; US-464499.
PR 23-AUG-1990; US-569781.
PA (DUPO ) DU FONT DE NEMOURS CO.
PI Dean C, Harder PA, Leto KU, Lichtner FT, Odell JT;
```

```
PI O'Keefe DP, Omer CA, Romesser JA;
DR WPI; 91-102077/14.
DR N-PSDB; Q11126.
PT DNA encoding cytochrome P450 enzymes - and electron donating
PT iron-sulphur proteins, used to confer herbicide resistance to
PT plants and microorganisms
PS Claim 13; page 151; 224pp; English.
CC This cytochrome P450 enzyme, P450SU1 is expressed alongside the iron
CC sulphur protein P450-B, by a DNA sequence contained in a recombinant
CC plasmid. Host Streptomyces species are transformed with the plasmid
CC and are used to coat a plant seed to transform the plant. The res-
CC ultant transformants are resistant to herbicides.
CC See also Q11127.
SQ Sequence 406 AA;

Query Match 13.9%; Score 62.5; DB 1; Length 406;
Best Local Similarity 35.5%; Pred. No. 4.8;
Matches 22; Conservative 9; Mismatches 28; Indels 3; Gaps 3;

QY 19 DAIVADNPR-AVVVESP-GMVKIDAPDRLTIRRETIEELTGTFRDLQLOVNLITLSGHI 76
Db 76 DNFFATSPRFAVRESPOAFGLDPPEHGTTRRTISETVKRIKGMPEVEEV-VHGFL 134
QY 77 DE 78
Db 135 DE 136

RESULT 14
R61118
ID R61118 standard; Protein; 322 AA.
AC R61118;
DE 07-APR-1995 (first entry)
DE Yeast presequence COX IV-wheat COX II (unedited) fusion protein.
KW Yeast cytochrome oxidase subunit IV presequence; cox IV; cox II;
KW wheat cytochrome oxidase subunit II; chimeric gene;
KW unedited mitochondrial gene; transgenic plant; male-sterile plant;
KW transgene; fusion protein.
OS Chimeric Saccharomyces cerevisiae.
OS Chimeric Triticum sp.
FH Key Location/Qualifiers
FT peptide 1..62
FT /note= "yeast COX IV amino acids 1-62"
FT protein 63..322
FT /label= COX II
FT /note= "unedited version"
PN W09418334-A.
PD 18-AUG-1994.
PF 15-FEB-1994; F00162.
PR 15-FEB-1993; FR-001650.
PA (CNRS ) CNRS CENT NAT RECH SCI.
PI Araya A, Mouras A;
DR WPI; 94-279755/34.
DR N-PSDB; Q70275.
PT Transgenic plant contg non-edited mitochondrial gene - linked to
PT transfer sequence, able to impart male sterility without altering
PT other characteristics, also antisense constructs for restoring
PT fertility
PS Claim 4; Page 32-33; 64pp; French.
CC The COX II from wheat is fused to codons 1-62 of the subunit IV of
CC yeast cytochrome oxidase. The COX IV transfer sequence directs the
CC COX II to mitochondria. "Editing" the COX II gene results in
CC changes to 16 amino acids, compared to the unedited version. The
CC unedited COX II is preferred because it results in male-sterility
CC by inhibiting pollen production in transgenic plants.
SQ Sequence 322 AA;

Query Match 13.7%; Score 62; DB 1; Length 322;
Best Local Similarity 32.1%; Pred. No. 4.1;
Matches 27; Conservative 12; Mismatches 25; Indels 20; Gaps 6;
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:30:02 ; Search time 58.11 seconds

(without alignments)  
47.462 Million cell updates/sec

Title: US-09-430-029-4  
Perfect score: 451

Sequence: 1 MSNVFIAFOANEDSRPIVDA.....ITISGHIDEDEFTLSWSH 89

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	247	54.8	90	1 DMPM_PSESP	P19731 pseudomonas
2	122	27.1	102	1 TMOD_PSESP	Q00459 pseudomonas
3	69	15.3	822	1 MUS2_SYNY3	P73625 synechocyst
4	66.5	14.7	2481	1 UN52_CAEEL	Q06561 caenorhabdi
5	66	14.6	138	1 MMOB_METTR	P27356 methylosinu
6	66	14.6	181	1 CHMU_ERWHE	P42517 erwinia her
7	66	14.6	619	1 ELFL_HUMAN	P32519 homo sapien
8	65.5	14.5	392	1 METK_MESCR	P93254 mesembryant
9	65	14.4	1162	1 VGL2_IBV6	P12651 avian infec
10	65	14.4	1232	1 Y005_CAEEL	P34643 caenorhabdi
11	64.5	14.3	316	1 CORA_ECOLI	P27841 escherichia
12	64.5	14.3	452	1 KICO_MOUSE	Q61414 mus musculu
13	64	14.2	1162	1 VGL2_IBV6	P11223 avian infec
14	63.5	14.1	224	1 YHIC_BACSU	O05494 bacillus su
15	63.5	14.1	394	1 METK_HORVU	P50299 hordeum vul
16	63.5	14.1	420	1 DODA_ECOLI	P00861 escherichia
17	63	14.0	366	1 METK_PEA	P43612 pisum sativ
18	63	14.0	828	1 YEA4_YEAST	P43585 saccharomyc
19	62.5	13.9	205	1 CMEL_BACSU	P39694 bacillus su
20	62.5	13.9	405	1 CXPE_STRGO	P18326 streptomyce
21	62.5	13.9	964	1 Y0Y1_CAEEL	Q09560 caenorhabdi
22	62	13.7	85	1 Y00G_BPT4	P31285 bacterioph
23	62	13.7	145	1 METL_PETCR	P47916 petroselinu
24	62	13.7	395	1 METK_POPDE	Q60775 mus musculu
25	62	13.7	612	1 ELFL_MOUSE	P32927 homo sapien
26	62	13.7	897	1 CRFB_HUMAN	P53520 proteus mir
27	61.5	13.6	254	1 PMFD_PROMI	P47465 mycoplasma
28	61.5	13.6	411	1 Y223_LYCGE	P43281 lycopersico
29	61	13.5	393	1 METL_LYCGE	Q44145 anabaena sp
30	61	13.5	444	1 N1FN_ANASP	P28598 bacillus su
31	61	13.5	543	1 CH60_BACSU	P55624 rhizobium s
32	61	13.5	583	1 Y4OC_RHISN	P38351 saccharomyc
33	60.5	13.4	445	1 PAFL_YEAST	

## RESULT 1

ID	DMPM_PSESP	STANDARD;	PRT;	90 AA.
AC	P19731;			
DT	01-FEB-1991 (Rel. 17, Created)			
DT	01-FEB-1991 (Rel. 17, Last sequence update)			
DT	15-FEB-2000 (Rel. 39, Last annotation update)			
DE	PHENOL HYDROXYLASE P2 PROTEIN (EC 1.14.13.7) (PHENOL 2-MONOOXYGENASE P2 COMPONENT).			
DE	P2 COMPONENT).			
GN	DMPM OR PHEA3			
OS	Pseudomonas sp. (strain CF600).			
OG	Plasmid pVII50.			
OC	Bacteria; Proteobacteria.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE: 91072230.			
RA	Nordlund I., Powlowski J., Shingler V.;			
RT	*complete nucleotide sequence and polypeptide analysis of			
RT	multicomponent phenol hydroxylase from Pseudomonas sp. strain			
RT	CF600.*;			
RL	J. Bacteriol. 172:6826-6833(1990).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BH;			
RA	Takeo M., Maeda Y., Okada H., Miyama K., Mori K., Ike M.,			
RA	Fujita M.;			
RL	Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	STRUCTURE BY NMR.			
RX	MEDLINE: 97164903.			
RA	Qian H., Edlund U., Powlowski J., Shingler V., Sethson I.;			
RT	*Solution structure of phenol hydroxylase protein component P2			
RT	determined by NMR spectroscopy.*;			
RT	Biochemistry 36:495-504(1997).			
CC	!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED			
CC	DERIVATIVES. P2 IS REQUIRED FOR GROWTH ON PHENOL, AND FOR			
CC	IN VITRO PHENOL HYDROXYLASE ACTIVITY.			
CC	!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+)			
CC	+ H(2)O.			
CC	!- COFACTOR: FAD, AND REQUIRES FE(+2) FOR ACTIVITY.			
CC	!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.			
CC	!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED			
CC	BY P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
CC	EMBL; M60276; AAA25941.1; -			
DR	EMBL; D28864; BAA06016.1; -			
DR	PIR; C37831; C37831.			
DR	PDB; 1HQI; 23-DEC-96.			

## ALIGNMENTS

34	60.5	13.4	1050	1	ULK1_HUMAN	O75385	homo sapien
35	60.5	13.4	1154	1	VGL2_IBVD2	P12722	avian infec
36	60.5	13.4	1163	1	VGL2_IBV6	P05135	avian infec
37	59.5	13.2	597	1	METK_STAAU	P50307	staphylococ
38	59.5	13.2	748	1	TREE_EMENI	O42777	emeritella
39	59	13.1	234	1	METK_PPTCR	P31155	petroselinu
40	59	13.1	376	1	KAPR_CAEEL	P30625	caenorhabdi
41	59	13.1	385	1	SUCC_AQUAE	O67546	aquifex aeo
42	59	13.1	421	1	FTSZ_HAEIN	P45069	haemophilus
43	59	13.1	1423	1	FRUA_STRMU	Q03174	streptococc
44	58.5	13.0	298	1	CBBR_RHORU	P52595	rhodospiril
45	58.5	13.0	393	1	METK_MUSAC	O22338	musa acumin

```
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
SQ NADP; Flavoprotein; FAD; Iron; Plasmid; 3D-structure.
KW SEQUENCE 90 AA; 10491 MW; 32B3A5FB72664AED CRC64;

Query Match 54.8%; Score 247; DB 1; Length 90;
Best Local Similarity 51.1%; Pred. No. 5.3e-20;
Matches 45; Conservative 21; Mismatches 22; Indels 0; Gaps 0;

Qy 2 SNVFTAFQANEDSRPIVDAIVADNPRAVVVSPGMVKIDAPDLRTIRRETIETELGTFRD 61
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 SLVYTAFOQDNARVYVETIQDNPVAVVQHHPAMIRTEAEKRLREIRRETVENILGRWD 62
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 62 LQQLQVNLITLSGHIDEDEFTLSWSH 89
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 63 VQEMLVDTITGGVDEDDDFVLEKN 90
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
TMOB_PSEME STANDARD; PRT; 102 AA.
ID Q00459;
AC Q00459;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TOLUENE-4-MONOOXYGENASE SYSTEM PROTEIN D (EC 1.14.13.-).
GN TMOB.
OS Pseudomonas mendocina.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-21.
RC STRAIN-KR1.
RX MEDLINE; 91358306.
RA Yen K.-M., Karl M.R., Blatt L.M., Simon M.J., Winter R.B.,
RA Fausset P.R., Lu H.S., Harcourt A.A., Chen K.K.;
RA "Cloning and characterization of a Pseudomonas mendocina KRL gene
RT cluster encoding toluene-4-monooxygenase.";
RL J. Bacteriol. 173:5315-5327(1991).
CC -!- FUNCTION: HYDROXYLATES TOLUENE TO FORM P-CRESOL.
CC -!- COFACTOR: FAD; REQUIRES FE(+2) FOR ACTIVITY.
CC -!- PATHWAY: FIRST STEP IN TOLUENE DEGRADATION.
CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME TOLUENE-4-MONOOXYGENASE
CC IS FORMED BY THE TMOA, TMOB, TMOG, TMOH, TMOI, TMOJ AND TMOK
CC POLYPEPTIDES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M65106; AAA26002.1; -.
DR HSSP; P19731; 1HQ1.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Flavoprotein;
KW Monooxygenase; FAD; Iron.
FT INIT_MET 0
SQ SEQUENCE 102 AA; 11487 MW; E5E4283F8E2904B3 CRC64;

Query Match 27.1%; Score 122; DB 1; Length 102;
Best Local Similarity 32.5%; Pred. No. 1.5e-06;
Matches 27; Conservative 21; Mismatches 33; Indels 2; Gaps 2;

Qy 2 SNVFTAFQANEDSRPIVDAIVADNPRAVVVSPGMVKIDAPDLRTIRRETIETELGTGR 59
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 11 NVNGVPIIRAGDLVEVETAEIDNPGKEITVDHRYVRIAEAGELIILTRKTLEQLGRP 70
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 60 FDLQQLVNLITLSGHIDEDE 82
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71 FNNQLEINLASFAGQIQADEQ 93
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

us-09-430-029-4.rsp
```

```
RESULT 3
MUS2_SINY3 STANDARD; PRT; 822 AA.
ID MUS2_SINY3
AC P73625;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE MUTS2 PROTEIN.
GN MUTSB OR MUTS2 OR SLL1772.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
RN SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- FUNCTION: NOT KNOWN.
CC -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90908; BAAL1670.1; -.
DR PFAM; PF00488; Muts_C; 1.
DR PFAM; PF01713; Smr; 1.
DR PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; FALSE_NEG.
KW ATP-binding; DNA-binding.
FT NP_BIND 348 355 ATP (POTENTIAL).
SQ SEQUENCE 822 AA; 90669 MW; 78AE7942F77F721 CRC64;

Query Match 15.3%; Score 69; DB 1; Length 822;
Best Local Similarity 36.1%; Pred. No. 8.6;
Matches 26; Conservative 11; Mismatches 27; Indels 8; Gaps 3;

Qy 13 DSRPIVDAIVADNPRAVVVSPGMVKIDAPDLRTIRRETIETELGTFRDLOQ 64
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 338 DSQIRVIAITGPTGKTVTKTIGLVALMAKVGLYIPAKEIVEMFWFAIILADIGDEQS 397
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 65 LQVNLITLSGHI 76
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 398 LQVNLITLSGHI 409
   1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
UN52_CAEL STANDARD; PRT; 2481 AA.
ID UN52_CAEL
AC Q06561;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE BASEMENT MEMBRANE PROTEOLYCAN PRECURSOR (PERLECAN HOMOLOG).
GN UNC-52.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
RN SEQUENCE FROM N.A.
RX MEDLINE; 9339574.

Page 2
```

RA Rogalski T.M., Williams B.D., Mullen G.P., Moerman D.G.;  
 RT "Products of the unc-52 gene in *Caenorhabditis elegans* are homologous  
 to the core protein of the mammalian basement membrane heparan  
 sulfate proteoglycan";  
 RL Genes Dev. 7:1471-1484(1993).  
 CC -1- FUNCTION: PROBABLE ROLE IN MYOFILAMENT ASSEMBLY AND/OR ATTACHMENT  
 OF THE MYOFILAMENT LATTICE TO THE CELL MEMBRANE. UNC-52 MAY BE AN  
 CC EXTRACELLULAR ANCHOR FOR INTEGRIN RECEPTORS IN MUSCLE.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.  
 CC -1- ALTERNATIVE PRODUCTS: UNC-52 PRODUCES AT LEAST THREE POLYPEPTIDES:  
 CC ONE VERY SHORT FORM AND TWO LONG FORMS.  
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANE OF ALL  
 CC CONTRACTILE TISSUES. IT IS CONCENTRATED OVER MUSCLE DENSE BODIES  
 CC AND M-LINES WHICH ARE ASSOCIATED WITH BETA-INTEGRIN.  
 CC -1- DEVELOPMENTAL STAGE: SYNTHESIZED EARLY IN EMBRYOGENESIS.  
 CC -1- SIMILARITY: CONTAINS 3 LDL-RECEPTOR CLASS A DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 16 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 7 LAMININ EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.  
 CC -1- SIMILARITY: TO PERLECAN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L13458; AAA28156.1; --  
 DR HSP; P01130; IAJJ.  
 DR PFAM; PF00047; ig; 16.  
 DR PFAM; PF00052; laminin\_B; 2.  
 DR PFAM; PF00053; laminin\_EGF; 5.  
 DR PFAM; PF00057; ldl\_recept.a; 3.  
 DR PROSITE; PS00022; EGF\_1; 4.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS01209; LDLRA\_1; 3.  
 DR PROSITE; PS01209; LDLRA\_2; 3.  
 DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 7.  
 KW Proteoglycan; Glycoprotein; Signal; Immunoglobulin domain; Repeat;  
 KW Basement membrane; Extracellular matrix; Alternative splicing;  
 KW Laminin EGF-like domain.  
 FT SIGNAL 1 18  
 FT CHAIN 19 2481  
 FT DOMAIN 19 148  
 FT DOMAIN 149 188  
 FT DOMAIN 190 229  
 FT DOMAIN 233 272  
 FT DOMAIN 273 363  
 FT DOMAIN 384 431  
 FT DOMAIN 432 441  
 FT DOMAIN 442 632  
 FT DOMAIN 633 665  
 FT DOMAIN 673 719  
 FT DOMAIN 720 729  
 FT DOMAIN 730 920  
 FT DOMAIN 921 953  
 FT DOMAIN 954 1003  
 FT DOMAIN 1010 1059  
 FT DOMAIN 1060 1110  
 FT DOMAIN 1133 1216  
 FT DOMAIN 1227 1311  
 FT DOMAIN 1318 1403  
 FT DOMAIN 1415 1499  
 FT DOMAIN 1507 1592  
 FT DOMAIN 1598 1682  
 FT DOMAIN 1699 1785  
 FT DOMAIN 1794 1878  
 FT DOMAIN 1887 1976  
 FT DOMAIN 1977 2070  
 FT DOMAIN 2078 2162  
 FT DOMAIN 2174 2260

DOMAIN 2267 2347  
 FT DOMAIN 2354 2436  
 FT DISULFID 66 114  
 FT DISULFID 149 161  
 FT DISULFID 156 174  
 FT DISULFID 168 183  
 FT DISULFID 190 202  
 FT DISULFID 197 215  
 FT DISULFID 209 224  
 FT DISULFID 233 246  
 FT DISULFID 240 259  
 FT DISULFID 253 268  
 FT DISULFID 954 963  
 FT DISULFID 956 970  
 FT DISULFID 973 982  
 FT DISULFID 985 1001  
 FT DISULFID 1010 1020  
 FT DISULFID 1012 1026  
 FT DISULFID 1029 1038  
 FT DISULFID 1041 1057  
 FT DISULFID 1060 1068  
 FT DISULFID 1062 1078  
 FT DISULFID 1081 1090  
 FT DISULFID 1093 1108  
 FT DISULFID 1151 1199  
 FT DISULFID 1337 1383  
 FT DISULFID 1434 1480  
 FT DISULFID 1526 1572  
 FT DISULFID 1617 1662  
 FT DISULFID 1718 1766  
 FT DISULFID 1813 1860  
 FT DISULFID 1906 1953  
 FT DISULFID 1997 2052  
 FT DISULFID 2098 2146  
 FT DISULFID 2194 2241  
 FT DISULFID 2283 2328  
 FT DISULFID 2373 2419  
 FT CARBOHYD 1421 1421  
 SQ SEQUENCE 2481 AA; 270812 MW; E47FBE2B310F5031 CRC64;  
 Query Match 14.7%; Score 66.5; DB 1; Length 2481;  
 Best Local Similarity 28.4%; Pred. NO. 60;  
 Matches 21; Conservative 14; Mismatches 24; Indels 15; Gaps 3;  
 QY 16 PIVDAIVADNPRAVYVVEPGMKVIDAPDELITIRRETIETLTGTREDLQ-----QLQVN 68  
 Db 1577 PTTNRPVSNPRAVIVKSPIRPIDPAE-----QTVPE--GSPFKIRCYVPGHPSVOLT 1628  
 QY 69 LITLSGHIDEDDDE 82  
 Db 1629 FRRVSSQLNEDADE 1642  
 RESULT 5  
 MMOV METTR STANDARD; PRT; 138 AA.  
 AC P27356;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DE 01-AUG-1992 (Rel. 23, Last annotation update)  
 DE METHANE MONOOXYGENASE REGULATORY PROTEIN B.  
 GN MMOV.  
 OS Methylosinus trichosporium.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Methylocystaceae; Methylosinus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 91251762.  
 RA Cardy D.L.N., Laidler V., Salmond G.P.C., Murrell J.C.;  
 RT "Molecular analysis of the methane monooxygenase (MMO) gene cluster  
 of *Methylosinus trichosporium* OB3b.";  
 RL Mol. Microbiol. 5:335-342(1991).

CC -1- FUNCTION: THE B PROTEIN ACTS AS A REGULATOR OF ELECTRON FLOW  
 CC THROUGH THE SOLUBLE MMO COMPLEX, SWITCHING THE ENZYME FROM AN  
 CC OXIDASE TO A HYDROXYLASE IN THE PRESENCE OF THE SUBSTRATE.  
 CC -1- SUBUNIT: M. TRICHOSPORUM HAS TWO FORMS OF METHANE MONOOXYGENASE,  
 CC A SOLUBLE AND A MEMBRANE-BOUND TYPE. THE SOLUBLE TYPE CONSISTS  
 CC OF THREE COMPONENTS (A, B AND C).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X55394; CAA39070.1; -  
 DR PIR: S15209; S15209.  
 KW Oxidoreductase; Monooxygenase.  
 SQ SEQUENCE 138 AA; 14883 MW; B8FB8731DF525E82 CRC64;  
 Query Match 14.6%; Score 66; DB 1; Length 138;  
 Best Local Similarity 25.5%; Pred. No. 2.1;  
 Matches 24; Conservative 12; Mismatches 46; Indels 12; Gaps 2;  
 QY 2 SNVFIAFQANEDSRPIYDAIVAD-----NPRVVVSPGMVKIDAPDLIRIRRETI 52  
 Db 35 SNAVVLVLMKSDSDE--DAIIEDIVLKGKAKNPSIVVEDKAGFWIKADGAIEIDAAEA 91  
 QY 53 EELTGTRFDLQQLVNLTLSGHIDEDEDEFTLS 86  
 Db 92 GELGKGFVYDLLINVSSTVGAYTIGTKPTIT 125  
 RESULT 6  
 CHMO\_ERWHE  
 ID CHMO\_ERWHE STANDARD; PRT; 181 AA.  
 AC P42517;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE MONOFUNCTIONAL CHORISMATE MUTASE PRECURSOR (EC 5.4.99.5) (CM-F).  
 GN AROQ.  
 OS Erwinia herbicola.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Erwinia.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-34.  
 RC STRAIN=ATCC 33243;  
 RX MEDLINE; 93328677.  
 RA Xia T., Song J., Zhao G., Aldrich H., Jensen R.A.;  
 RT "The aroQ-encoded monofunctional chorismate mutase (CM-F) protein is  
 RT a periplasmic enzyme in Erwinia herbicola."  
 RL J. Bacteriol. 175:4729-4737(1993).  
 CC -1- FUNCTION: MAY SEQUESTER WITH CYCLOHEXADIENYL DEHYDRATASE AND  
 CC PERHAPS AN AMINOTRANSFERASE TO FORM PHENYLALANINE OR  
 CC PHENYLPIRUVATE FOR SOME AS YET UNKNOWN FUNCTION.  
 CC -1- CATALYTIC ACTIVITY: CHORISMATE = PREPHENATE.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M95628; AAA73360.1; -  
 DR PIR: PF01817; Chorismate.mut; 1.  
 KW Isomerase; Periplasmic; Signal.  
 FT SIGNAL 1 20

FT CHAIN 21 181 MONOFUNCTIONAL CHORISMATE MUTASE.  
 SQ SEQUENCE 181 AA; 20299 MW; 5F8F77C0D1E2542E CRC64;  
 Query Match 14.6%; Score 66; DB 1; Length 181;  
 Best Local Similarity 29.6%; Pred. No. 2.9;  
 Matches 16; Conservative 11; Mismatches 17; Indels 10; Gaps 2;  
 QY 39 IDAPDLIRIRETEELTGTRFDLQQLVNLT-----LSGHIDEDEDEFTLS 86  
 Db 106 LSSPDSAV-----PVRDLTETTRQIQQLDTQLLTALISQRLMTGAFSQDEKEFLMS 155  
 RESULT 7  
 ELF1\_HUMAN  
 ID ELF1\_HUMAN STANDARD; PRT; 619 AA.  
 AC P32519;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE ETS-RELATED TRANSCRIPTION FACTOR ELF-1 (E74-LIKE FACTOR 1).  
 GN ELF1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 92407982.  
 RA Leiden J.M., Wang C.Y., Petryniak B., Markovitz D.M., Nabel G.J.,  
 RA Thompson C.B.;  
 RT "A novel Ets-related transcription factor, Elf-1, binds to human  
 RT immunodeficiency virus type 2 regulatory elements that are required  
 RT for inducible trans activation in T cells."  
 RL J. Virol. 66:5890-5897(1992).  
 RN [2]  
 RP BINDING TO RB.  
 RX MEDLINE; 93262492.  
 RA Wang C.Y., Petryniak B., Thompson C.B., Kaelin W.G., Leiden J.M.;  
 RT "Regulation of the Ets-related transcription factor Elf-1 by binding  
 RT to the retinoblastoma protein."  
 RL Science 260:1330-1335(1993).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT APPEARS TO BE REQUIRED FOR THE  
 CC T-CELL-RECEPTOR-MEDIATED TRANS ACTIVATION OF HIV-2 GENE  
 CC EXPRESSION. BINDS SPECIFICALLY TO TWO PURINE-RICH MOTIFS IN THE  
 CC HIV-2 ENHANCER. ELF-1 BINDS TO THE UNDERPHOSPHORYLATED FORM OF RB.  
 CC MAY INTERACT WITH OTHER TRANSCRIPTION FACTORS IN ORDER TO REGULATE  
 CC SPECIFIC GENES.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M82882; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: A43361; A43361.  
 DR HSSP; P14921; 2STW.  
 DR TRANSFAC; T01113; -.  
 DR MIM; 189973; -.  
 DR PRINTS; PF00178; Ets; 1.  
 DR PRAM; PR00454; ETSDOMAIN  
 DR PROSITE; PS00345; ETS\_DOMAIN\_1; 1.  
 DR PROSITE; PS00346; ETS\_DOMAIN\_2; 1.  
 DR PROSITE; PS00061; ETS\_DOMAIN\_3; 1.  
 KW Nuclear protein; Transcription regulation; Activator; DNA-binding.  
 FT DOMAIN 75 80 POLY-ASP.  
 FT DNA\_BIND 208 290  
 FT SEQUENCE 619 AA; 67455 MW; AB0B41B2964A66EE CRC64;

```

Query Match          14.6%; Score 66; DB 1; Length 619;
Best Local Similarity 24.1%; Pred. No. 13;
Matches 21; Conservative 20; Mismatches 34; Indels 12; Gaps 3;

QY 2 SNVTFIAQAN--EDSRPIVDAIVADNPRAVVVSPGWMKIDAPDRLTTERETIELTGTGR 59
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 NDLVFEFASNYWDERQLGDRPAIF--PAVIVHEVPGADILNSYAGLACVEFPNDMITSS 65
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 FDLOQLQVNLITLSCHIDEDDEFTLS 86
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 LDVAEEI-----IDDDDDITLT 84
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
METK_MESCR
ID METK_MESCR STANDARD; PRT; 392 AA.
AC P93254;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE S-ADENOSYLMETHIONINE SYNTHETASE (EC 2.5.1.6) (METHIONINE
DE ADENOSYLTRANSFERASE) (ADOMET SYNTHETASE).
OS Mesembryanthemum crystallinum (Common ice plant).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Caryophyllales; Caryophyllales;
OC Aizoaceae; Mesembryanthemum.
RN [1]
RP SEQUENCE FROM N.A.
RA Michalowski C.B., Bohnert H.J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM
CC METHIONINE AND ATP (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + L-METHIONINE + H(2)O = ORTHOPHOSPHATE +
CC PYROPHOSPHATE + S-ADENOSYL-L-METHIONINE.
CC -!- PATHWAY: ACTIVATED METHYL CYCLE.
CC -!- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U79767; AAB38500.1; -
CC DR HSP; P04384; 1XRC
CC DR PFAM; PF00438; S-AdoMet_synt; 1.
CC DR PROSITE; PS00376; ADOMET_SYNTHETASE_1; 1.
CC DR PROSITE; PS00377; ADOMET_SYNTHETASE_2; 1.
CC KW Transferase; One-carbon metabolism; Multigene family; ATP-binding.
CC NP_BIND 119 124 ATP (POTENTIAL).
CC FT BINDING 147 147 ATP (POTENTIAL).
CC FT BINDING 147 147 ATP (POTENTIAL).
CC SEQUENCE 392 AA; 42885 MW; D13C2CD03757165A CRC64;

Query Match          14.5%; Score 65.5; DB 1; Length 392;
Best Local Similarity 29.2%; Pred. No. 8.4;
Matches 33; Conservative 9; Mismatches 38; Indels 33; Gaps 5;

QY 2 SNVTFIAQANE-----SRPIVD--AIVADNPRAVVVSPGWMKIDAPDRLTIRRET 51
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 281 SGVIAIQAASIVAAAGLARCIVQISVAGVAPLPSVFVDYTGTKISKDKILKIVKET 340
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 52 IEELTGTFRDLOQLQVNLITLS-----GHIDEDDEFT-----LSW 87
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 FDFRPG-----MTANLLDLRGGRNYLKTAAYAGHFGGRDADFTWEVVKPLK 387
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
VGL2_IBVM
ID VGL2_IBVM STANDARD; PRT; 1162 AA.

```

```

AC P12651;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN)
DE [CONTAINS: SPIKE PROTEIN S1; SPIKE PROTEIN S2].
GN S.
OS Avian infectious bronchitis virus (strain M41) (IBV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
RN [1]
RP SEQUENCE FROM N.A.
RA Nielse H.G.M., Lenstra J.A., Spaan W.J.M., Zijderveld A.J.,
RA Bleumink-Pluym N.M.C., Hong F., van Scharrenburg G.J.M.,
RA Horzinek M.C., van der Zeijst B.A.M.;
RT "The peplomer protein sequence of the M41 strain of coronavirus IBV
RT and its comparison with Beaudette strains.";
RL Virus Res. 5:253-263(1986).
CC -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M21883; AAA66575.1; -
CC DR EMBL; A24863; CAA01736.1; -
CC PIR; S07421; S07421.
CC PFAM; PF01600; Corona_S1; 1.
CC PFAM; PF01601; Corona_S2; 1.
CC KW Glycoprotein; Envelope protein; Transmembrane; Signal.
CC FT SIGNAL 1 18
CC FT CHAIN 19 1162 E2 GLYCOPROTEIN.
CC FT CHAIN 19 537 SPIKE PROTEIN S1.
CC FT CHAIN 538 1162 SPIKE PROTEIN S2.
CC FT DOMAIN 1120 1137 CYS-RICH.
CC FT CARBOHYD 51 51 POTENTIAL.
CC FT CARBOHYD 77 77 POTENTIAL.
CC FT CARBOHYD 103 103 POTENTIAL.
CC FT CARBOHYD 144 144 POTENTIAL.
CC FT CARBOHYD 163 163 POTENTIAL.
CC FT CARBOHYD 178 178 POTENTIAL.
CC FT CARBOHYD 212 212 POTENTIAL.
CC FT CARBOHYD 237 237 POTENTIAL.
CC FT CARBOHYD 247 247 POTENTIAL.
CC FT CARBOHYD 264 264 POTENTIAL.
CC FT CARBOHYD 271 271 POTENTIAL.
CC FT CARBOHYD 276 276 POTENTIAL.
CC FT CARBOHYD 306 306 POTENTIAL.
CC FT CARBOHYD 425 425 POTENTIAL.
CC FT CARBOHYD 447 447 POTENTIAL.
CC FT CARBOHYD 513 513 POTENTIAL.
CC FT CARBOHYD 530 530 POTENTIAL.
CC FT CARBOHYD 579 579 POTENTIAL.
CC FT CARBOHYD 591 591 POTENTIAL.
CC FT CARBOHYD 669 669 POTENTIAL.
CC FT CARBOHYD 676 676 POTENTIAL.
CC FT CARBOHYD 714 714 POTENTIAL.
CC FT CARBOHYD 947 947 POTENTIAL.
CC FT CARBOHYD 960 960 POTENTIAL.
CC FT CARBOHYD 979 979 POTENTIAL.
CC FT CARBOHYD 1014 1014 POTENTIAL.
CC FT CARBOHYD 1038 1038 POTENTIAL.
CC FT CARBOHYD 1051 1051 POTENTIAL.
CC FT CARBOHYD 1074 1074 POTENTIAL.
CC SEQUENCE 1162 AA; 128077 MW; 3C9CC70938492DDA CRC64;

```

```

Query Match      14.4%; Score 65; DB 1; Length 1162;
Best Local Similarity 26.7%; Pred. No. 35;
Matches 28; Conservative 11; Mismatches 46; Indels 20; Gaps 3;

QY 4 VETAFQANEDSRPIVDAIVA-----DNPRVAVVSPGMVKIDAPDRITIR 48
||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 934 VFHFSTPDSEFVNVTAVGFCVKPANASQYAIVPANGRGIFIQVNGSYITATDMYMPR 993

QY 49 RETIEEL-TGTRFDLOQVNLITLSGHIDDD---DEFTLSWS 88
| : : | | | : : | | : : | | | | | | | | | | | | | | | |
Db 994 AITAGDIVLTISQANVSVNKNVTITTFVNDDFDFNDELSKWN 1038

RESULT 10
YQ05_CAEEL
ID_YQ05_CAEEL STANDARD; PRT; 1232 AA.
AC P34643;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOTHETICAL 134.9 KDA PROTEIN ZK512.5 IN CHROMOSOME III.
ZK512.5
OS Caenorhabditis elegans.
OC Rhabditidae; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae;
OC Rhabditidae; Peloderidae; Caenorhabditis.
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22177; CAA80146.1; -.
DR PIR; S40766; S40766.
DR WORMPEP; ZK512.5; CE00411.
KW Hypothetical protein.
SQ SEQUENCE 1232 AA; 134923 MW; 6DFC35D664AA8D6A CRC64;

Query Match      14.4%; Score 65; DB 1; Length 1232;
Best Local Similarity 25.3%; Pred. No. 37;
Matches 19; Conservative 17; Mismatches 23; Indels 16; Gaps 3;

QY 17 IVDAIVADNPRAVAVVSPGMVKIDAPDRITIRITIEL-----TGTRFDLO 63
: ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 472 VVDALLVQLLEIMVQOHG--RVTPGVDAVLTITNASELGEKGTISSNGTSESGSKFEAK 529

QY 64 QLQVNLITLSGHIDE 78
: : : | | | | |
Db 530 E-RENKYLGGHINE 543

RESULT 11

```

```

CORAECOLI
ID CORAECOLI STANDARD; PRT; 316 AA.
AC P27841;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MAGNESIUM AND COBALT TRANSPORT PROTEIN CORA.
GN CORA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE; 93300795.
RA Smith R.L., Banks J.L., Snavely M.D., Maguire M.E.;
RT "Sequence and topology of the Cora magnesium transport systems of
RT Salmonella typhimurium and Escherichia coli. Identification of a new
RT class of transport protein."
RL J. Biol. Chem. 268:14071-14080(1993).
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12;
PC Ohmori H.;
RA Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 92358234.
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
RT from 84.5 to 86.5 minutes."
RL Science 257:771-778(1992).
[4]
RN RP REVISIONS.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 93347969.
RA Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;
RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT region from 87.2 to 89.2 minutes."
RL Nucleic Acids Res. 21:3391-3398(1993).
CC -1- FUNCTION: PLAYS A ROLE IN THE TRANSPORT OF MAGNESIUM AND COBALT
CC IONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L11042; AAB59046.1; -.
DR EMBL; L02122; AAD15038.1; -.
DR EMBL; M87049; AAA67612.1; -.
DR EMBL; AE000457; AAC76819.1; -.
DR PIR; S30743; S30743.
DR PIR; S30706; S30706.
DR ECOGENE; EG11463; CORA.
DR PFAM; PF01544; CorA; 1.
KW Magnesium; Cobalt; Transmembrane; Transport.
FT TRANSMEM 255 273 POTENTIAL.
FT TRANSMEM 288 310 POTENTIAL.
FT CONFLICT 49 50 SL -> RP (IN REF. 1 AND 3).
FT CONFLICT 303 303 G -> A (IN REF. 1).
SQ SEQUENCE 316 AA; 36589 MW; CC3B1B736EE36A53 CRC64;

Query Match      14.3%; Score 64.5; DB 1; Length 316;
Best Local Similarity 28.6%; Pred. No. 8.2;
Matches 20; Conservative 15; Mismatches 18; Indels 17; Gaps 3;

QY 12 EDSRPVDAIVADNPRAVAVVSPGMVKIDAPDRITIRITIELTGTRFDLOQLQVNLIT 71
| : : | | : : | | : : | | : : | | : : | | : : | | : : | |

```



```
Db 18 EESQPLNAVWID-----LVFPDDDERLRVQSELGQSL-ATRPELEDIEA----- 61
QY 72 LSGHIDEDD 81
Db 62 -SARFFEDD 70

RESULT 12
K1CO_MOUSE STANDARD; PRT; 452 AA.
AC Q61414;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE KERATIN, TYPE I CYTOSKELETAL 15 (CYTOKERATIN 15) (K15) (CK 15).
GN KRT15 OR KRT1-15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=LIVER;
RX MEDLINE; 94171037.
RA Nozaki M., Mori M., Matsushiro A.;
RT "The complete sequence of the gene encoding mouse cytokeratin 15.";
RL Gene 138:197-200(1994).
CC -1- FUNCTION: THERE ARE TWO TYPES OF CYTOSKELETAL AND MICROFIBRILLAR
CC KERATIN: I (ACIDIC; 40-55 KDA) [K9 TO K20] AND II (NEUTRAL TO
CC BASIC; 56-70 KDA) [K1 TO K8]. BOTH A BASIC AND AN ACIDIC KERATIN
CC ARE REQUIRED FOR FILAMENT ASSEMBLY.
CC -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D16313; BAA03821.1; -.
CC MGD; MGI:96689; KRT1-15.
CC PFAM; PF00038; filament; 1.
CC PROSITE; PS00226; IF; 1.
CC
CC Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin.
KW DOMAIN 1 97 HEAD.
FT DOMAIN 98 406 ROD.
FT DOMAIN 407 452 TAIL.
FT DOMAIN 98 133 COIL 1A.
FT DOMAIN 134 152 LINKER 1.
FT DOMAIN 153 244 COIL 1B.
FT DOMAIN 245 264 LINKER 12.
FT DOMAIN 265 406 COIL 2.
FT DOMAIN 36 39 POLY-GLY.
FT DOMAIN 421 424 POLY-GLY.
FT DOMAIN 425 430 POLY-SER.
SQ SEQUENCE 452 AA; 49159 MW; C431310A7BB408FD CRC64;

Query Match 14.3%; Score 64.5; DB 1; Length 452;
Best Local Similarity 26.0%; Pred. NO. 13;
Matches 25; Conservative 19; Mismatches 33; Indels 19; Gaps 4;

QY 8 FOANEDSRIVDAIVADNPRAW-VESPGMV-----KIDAPDRLTIR-----RETI 52
DB 150 KTIMEEIRDKILAAATIDRSRVLEIDNARLAADDFRLKYENELTLRQGVQVADINGLRRLV 209
QY 53 DELTGTFRDL----OOLQVNLITLSLGHIDEDDDEFT 84
DB 210 DELTLARTDLEMQLEQLNEELAYLAKKNHEEMKEFS 245
```

```
RESULT 13
VGL2_IBVB STANDARD; PRT; 1162 AA.
ID VGL2_IBVB
AC P11223; P05134;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN)
DE [CONTAINS: SPIKE PROTEIN S1; SPIKE PROTEIN S2].
GN S.
OS Avian infectious bronchitis virus (strain Beaudette) (IBV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85159540.
RA Binns M.M., Boursnell M.E.G., Cavanagh D., Pappind D.J.C.,
RA Brown T.D.K.;
RT "Cloning and sequencing of the gene encoding the spike protein of the
RT coronavirus IBV.";
RL J. Gen. Virol. 66:719-726(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87085499.
RA Binns M.M., Boursnell M.E.G., Tomley F.M., Brown D.K.;
RT "Comparison of the spike precursor sequences of coronavirus IBV
RT strains M41 and 6/82 with that of IBV Beaudette.";
RL J. Gen. Virol. 67:2825-2831(1986).
CC -1- FUNCTION: THE PELOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M95169; AAA70235.1; -.
CC EMBL; X02342; CAA26201.1; -.
CC DR PFAM; PF01600; Corona_S1; 1.
CC DR PFAM; PF01601; Corona_S2; 1.
CC
CC Glycoprotein; Envelope protein; Transmembrane; Signal.
KW SIGNAL 1 18
FT CHAIN 19 1162 E2 GLYCOPROTEIN.
FT CHAIN 19 537 SPIKE PROTEIN S1.
FT CHAIN 538 1162 SPIKE PROTEIN S2.
FT DOMAIN 1120 1137 CYS-RICH.
FT CARBOHYD 51 51 POTENTIAL.
FT CARBOHYD 77 77 POTENTIAL.
FT CARBOHYD 103 103 POTENTIAL.
FT CARBOHYD 144 144 POTENTIAL.
FT CARBOHYD 163 163 POTENTIAL.
FT CARBOHYD 178 178 POTENTIAL.
FT CARBOHYD 212 212 POTENTIAL.
FT CARBOHYD 237 237 POTENTIAL.
FT CARBOHYD 247 247 POTENTIAL.
FT CARBOHYD 264 264 POTENTIAL.
FT CARBOHYD 276 276 POTENTIAL.
FT CARBOHYD 306 306 POTENTIAL.
FT CARBOHYD 425 425 POTENTIAL.
FT CARBOHYD 447 447 POTENTIAL.
FT CARBOHYD 513 513 POTENTIAL.
FT CARBOHYD 530 530 POTENTIAL.
FT CARBOHYD 579 579 POTENTIAL.
FT CARBOHYD 591 591 POTENTIAL.
FT CARBOHYD 669 669 POTENTIAL.
FT CARBOHYD 676 676 POTENTIAL.
FT CARBOHYD 714 714 POTENTIAL.
FT CARBOHYD 947 947 POTENTIAL.
FT CARBOHYD 960 960 POTENTIAL.
FT CARBOHYD 979 979 POTENTIAL.
```

```

FT CARBOHYD 1014 1014 POTENTIAL.
FT CARBOHYD 1038 1038 POTENTIAL.
FT CARBOHYD 1051 1051 POTENTIAL.
FT CARBOHYD 1074 1074 POTENTIAL.
SQ SEQUENCE 1162 AA; 128046 MW; 0BAAD58113C9EBD5 CRC64;

Query Match 14.1%; Score 64; DB 1; Length 1162;
Best Local Similarity 26.7%; Pred. No. 45;
Matches 28; Conservative 11; Mismatches 46; Indels 20; Gaps 3;

QY 4 VFAFAQANESRPVDAIVA-----DNPRVVVSPGMVKIDAPDLRTIR 48
DQ 934 VFHFTVPSFVNVAIVGFCVKPANASQYAIVPANGRGIFQVNGSYITARDMTYMR 993
QY 49 RETI-EELTGRFDLOQLVNLTLSGHIDEDD---DEFTLSWS 88
DQ 994 AITAGDVTTTSCQANVSVNKVITVTFVNDNDFDNELSKWN 1038

RESULT 14
YDHC_BACSU STANDARD; PRT; 224 AA.
ID YDHC_BACSU STANDARD; PRT; 224 AA.
AC C05494;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHEICAL TRANSCRIPTIONAL REGULATOR IN DINE-PROB INTERGENIC REGION.
GN YDHC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=168 / MARBURG;
RA Sadaie Y., Yata K., Fujita M., Itaya M., Kasahara Y.,
RA Ogasawara N.;
RT "Nucleotide sequence and analysis of the phoB-rne-groESL region of
RT the Bacillus subtilis chromosome."
RL Microbiology 143:1861-1866(1997).
CC -1- SIMILARITY: BELONGS TO THE GNR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D88802; BAA19694.1; -
CC EMBL; Z99107; CAB12389.1; -
CC SUBTILIST; BG12180; YDHC.
CC PFAM; PF00392; gnr; 1.
CC PRINTS; PR00035; HTHGNTR.
CC PROSITE; PS00043; HTH_GNR_FAMILY; 1.
CC KW Hypothetical protein; Transcription regulation; DNA-binding.
CC DNA_BIND 37 56 H-T-H MOTIF (POTENTIAL).
CC SEQUENCE 224 AA; 26396 MW; 5C41A9E1553B592D CRC64;

Query Match 14.1%; Score 63.5; DB 1; Length 224;
Best Local Similarity 31.7%; Pred. No. 7;
Matches 19; Conservative 14; Mismatches 18; Indels 9; Gaps 3;

QY 30 VVSPGMVKIDAPDLRTIRRET---IHELGTFRFDLOQLVNLTLSGHIDE-DDDEFTL 85
DQ 57 LLEKDGILLKADRRNGFSITLTKADVDVEITYKIRIPLEQVAVELV-----IDEADEELTI 111

RESULT 15

```

```

METK_HORVU STANDARD; PRT; 394 AA.
ID METK_HORVU STANDARD; PRT; 394 AA.
AC P50299;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE S-ADENOSYLMETHIONINE SYNTHETASE 1 (EC 2.5.1.6) (METHIONINE
DE ADENOSYLTRANSFERASE 1) (ADOMET SYNTHETASE 1).
GN SAM1.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
RN [1]
RP SEQUENCE FROM N.A.
RA Mori S., Takizawa R.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE OF 285-322 FROM N.A.
RC TISSUE=ROOT;
RA Mori S., Takizawa R., Nakanishi H.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM
CC METHIONINE AND ATP.
CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + H(2)O = ORTHOPHOSPHATE +
CC PYROPHOSPHATE + S-ADENOSYL-L-METHIONINE.
CC -1- PATHWAY: ACTIVATED METHYL CYCLE.
CC -1- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D63835; BAA09895.1; -
CC EMBL; D49655; BAA08531.1; -
CC HSP; P04384; IXRC.
CC PFAM; PF00438; S-Adomet_synt; 1.
CC PROSITE; PS00376; ADOMET_SYNTHETASE_1; 1.
CC PROSITE; PS00377; ADOMET_SYNTHETASE_2; 1.
CC KW Transferase; One-carbon metabolism; Multigene family; ATP-binding.
CC NP_BIND 121 126 ATP (POTENTIAL).
CC BINDING 149 149 ATP (POTENTIAL).
CC SEQUENCE 394 AA; 42841 MW; 2ED786F33530013 CRC64;

Query Match 14.1%; Score 63.5; DB 1; Length 394;
Best Local Similarity 30.2%; Pred. No. 14;
Matches 32; Conservative 9; Mismatches 34; Indels 31; Gaps 6;

QY 2 SNVFIATQANED-----SRPIVD---AIVADNPRAVVPSPGMVKIDAPDLRTIRRET 51
DQ 283 SGAYIARQAASIIASGLARRCIIVSYAIGVPEPLSVFVDSYGTGKI--PDR-----E 334
QY 52 IEELTGTFRDLO--OLOVNL-----ITLSGHIDEDEFT 84
DQ 335 ILKLVKENDFRPGMITINLDLKKGNRFITAAAYGFGRRDADFT 380

Search completed: September 26, 2000, 20:30:04
Job time: 428 sec

```

2 . . . .

Wed Sep 27 17:17:02 2000

us-09-430-029-4.rsp

Page 9

-----

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:28:52 ; Search time 92.32 seconds  
(without alignments)  
66.841 Million cell updates/sec

Title: US-09-430-029-4  
Perfect score: 451  
Sequence: 1 MSNVFIAQANESRPVDA.....ITLSGHIDEDEFTLSWSH 89

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues  
Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_12:\*  
1: sp-archaea:\*  
2: sp-bacteria:\*  
3: sp-fungi:\*  
4: sp-human:\*  
5: sp-invertebrate:\*  
6: sp-mammal:\*  
7: sp-mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-virus:\*  
13: sp-vertebrate:\*  
14: sp-unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	354	78.5	89	Q30591	030591 burkholderi
2	346	76.7	89	Q25271	Q25271 pseudomonas
3	326	72.3	97	Q29NP5	Q29NP5 comamonas t
4	257	57.0	90	Q84960	Q84960 ralstonia s
5	247	54.8	90	Q52172	Q52172 pseudomonas
6	243	53.9	91	Q52163	Q52163 pseudomonas
7	233	51.7	89	Q32430	Q32430 acinetobact
8	232	51.4	89	Q43980	Q43980 acinetobact
9	114	25.3	110	Q27801	Q27801 pseudomonas
10	111	24.6	147	Q95411	Q95411 pseudomonas
11	110	24.4	104	Q51942	Q51942 burkholderi
12	105	23.3	105	Q69181	Q69181 alcaligenes
13	90	20.0	101	Q92ET4	Q92ET4 xanthobacte
14	84	18.6	104	Q07071	Q07071 burkholderi
15	83.5	18.5	862	Q29X8M6	Q29X8M6 streptomyce
16	77	17.1	1684	Q75218	Q75218 homo sapien
17	77	17.1	3830	Q9Y6H4	Q9Y6H4 homo sapien
18	77	17.1	3859	Q9Y631	Q9Y631 homo sapien
19	70.5	15.6	174	1	Q28973 archaeoglob

20	67.5	15.0	410	2	Q59831	streptomyce
21	66.5	14.7	2295	5	Q9XTD2	caenorhabdi
22	66.5	14.7	2482	5	O18263	caenorhabdi
23	66.5	14.7	3375	5	Q9XTI5	caenorhabdi
24	66	14.6	393	1	O30028	caenorhabdi
25	66	14.6	649	2	P73736	archaeoglob
26	65	14.4	263	5	O76870	synecocyst
27	65	14.4	281	2	Q9ZHS9	drosohila
28	65	14.4	1153	12	Q66178	streptomyce
29	65	14.4	1159	12	Q82624	avian infec
30	64	14.2	352	3	Q9Y887	suillus bov
31	64	14.2	395	4	Q9Y308	homo sapien
32	63.5	14.1	983	2	O30988	pseudomonas
33	63	14.0	138	2	O06118	methylocyst
34	63	14.0	557	2	Q00876	renibacteri
35	63	14.0	690	2	Q9X6R4	aeromonas p
36	62.5	13.9	139	1	O27443	methanobact
37	62.5	13.9	1541	2	O54298	streptomyce
38	62.5	13.9	4077	2	O52820	amycolatops
39	62	13.7	166	7	Q95517	macaca mula
40	62	13.7	274	8	Q37697	zea mays (m
41	62	13.7	354	7	Q95513	macaca fasc
42	62	13.7	354	7	Q95514	macaca fasc
43	62	13.7	811	12	O36185	plautia sta
44	62	13.7	941	11	O63128	rattus norv
45	62	13.7	1653	2	P76578	escherichia

## ALIGNMENTS

RESULT 1  
Q30591 PRELIMINARY; PRT; 89 AA.  
AC Q30591;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)  
DE CRPC PROTEIN.  
GN CRPC.  
OS Burkholderia pickettii (Pseudomonas pickettii).  
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;  
OC Ralstonia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PK01;  
RA OLSEN R.H., KUKOR J.J., BYRNE A.M., JOHNSON G.R.;  
RL J. Ind. Microbiol. 0:0-0(1997).  
DR EMBL; AF012632; AAB67107.1; -.  
DR HSSP; P19731; 1HQI.  
SQ SEQUENCE 89 AA; 10055 MW; BD74CF41 CRC32;

Query Match 78.5%; Score 354; DB 2; Length 89;  
Best Local Similarity 76.4%; Pred. No. 1.4e-29;  
Matches 68; Conservative 7; Mismatches 14; Indels 0; Gaps 0;  
QY 1 MSNVFIAQANESRPVDAIVADNPRAVVPVSGVMVKIDAPRLIRRETIETGTGRF 60  
Db 1 MSQVTFIAQANESRPVDAIVADNPRAVVPVSGVMVKIDAPRLIRRETIETGTGRF 60  
QY 61 DLQQLQVNLITLSGHIDEDEFTLSWSH 89  
Db 61 DLQQLHVNLTLSGHIDEDEFTLSWSH 89

RESULT 2  
Q25271 PRELIMINARY; PRT; 89 AA.  
AC Q25271;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

```

DE  TOLUENE/BENZENE-2-MONOXYGENASE (TBMA-TMMF) GENES, COMPLETE CDS'S
DE  (TBMA-TMMF).
GN  TBMC.
OS  Pseudomonas sp.
OC  Bacteria; Proteobacteria.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-JS150;
RX  MEDLINE; 96035667.
RA  JOHNSON G.R., OLSEN R.H.;
RT  "Nucleotide sequence analysis of genes encoding a toluene/benzene-2-
RL  monoxygenase from Pseudomonas sp. strain JS150.";
RL  Appl. Environ. Microbiol. 61:3336-3346(1995).
DR  EMBL; L40033; AAA88458.1; -.
DR  HSSP; P19731; IHQI.
SQ  SEQUENCE 89 AA; 10029 MW; F54D09A2 CRC32;

Query Match 76.7%; Score 346; DB 2; Length 89;
Best Local Similarity 75.3%; Pred. No. 9.3e-29;
Matches 67; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY  1 MSNVFIAGQNDSPRIYDAIVADNPRAVVPSPGMVKIDAPDRLTIRRETIETLTGTRF 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  1 MSQVFIAGQNDSPRIYDAIVADNPRAVVPSPGMVKIDAPDRLTIRRETIETLTGTRF 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  61 DLQQLQVNLITLSGHIDDDDEFTLSWS 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  61 DLQQLHVNLTLSGHIDDDDDQLTISWQH 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
Q9ZNP5 PRELIMINARY; PRT; 97 AA.
AC Q9ZNP5;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE COMPONENT.
GN APHM.
OS Comamonas testosteroni (Pseudomonas testosteroni).
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TA441;
RX MEDLINE; 99018839.
RA ARAI H., AKAHIRA S., OHISHI T., MAEDA M., KUDO T.;
RT "Adaptation of Comamonas testosteroni TA441 to utilize phenol:
RT organization and regulation of the genes involved in phenol
RT degradation.";
RL Microbiology 144:2895-2903(1998).
DR EMBL; AB006479; BAA34171.1; -.
DR HSSP; P19731; IHQI.
SQ SEQUENCE 97 AA; 10728 MW; 47CD2246 CRC32;

Query Match 72.3%; Score 326; DB 2; Length 97;
Best Local Similarity 69.3%; Pred. No. 1.2e-26;
Matches 61; Conservative 14; Mismatches 13; Indels 0; Gaps 0;

QY  1 MSNVFIAGQNDSPRIYDAIVADNPRAVVPSPGMVKIDAPDRLTIRRETIETLTGTRF 60
   :|||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  9 VSKVFIAGQNDSPRIYDAIVADNPRAVVPSPGMVKIDAPDRLTIRRETIETLTGTRF 68
   :|||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  61 DLQQLQVNLITLSGHIDDDDEFTLSWS 88
   :|||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  69 ELQQLQVNLITLSGHIDDDDDFTLSWN 96
   :|||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
O84960 PRELIMINARY; PRT; 90 AA.
ID O84960
AC O84960;

```

```

DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE COMPONENT.
GN POXC.
OS Ralstonia sp. E2.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Ralstonia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E2;
RX MEDLINE; 98361023.
RA HINO S., WATANABE K., TAKAHASHI N.;
RT "Phenol hydroxylase cloned from Ralstonia eutropha strain E2 exhibits
RL novel kinetic properties.";
RL Microbiology 144:0-0(0).
DR EMBL; AF026065; AAC32454.1; -.
DR HSSP; P19731; IHQI.
SQ SEQUENCE 90 AA; 9956 MW; 8B0F5684 CRC32;

Query Match 57.0%; Score 257; DB 2; Length 90;
Best Local Similarity 57.5%; Pred. No. 1.3e-19;
Matches 50; Conservative 16; Mismatches 21; Indels 0; Gaps 0;

QY  2 SNVFIAGQNDSPRIYDAIVADNPRAVVPSPGMVKIDAPDRLTIRRETIETLTGTRF 61
   :|||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  3 ANVYIALQNDSPRIYDAIVADNPRAVVPSPGMVKIDAPDRLTIRRETIETLTGTRF 62
   :|||: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  62 LQQLQVNLITLSGHIDDDDEFTLSWS 88
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB  63 LQEIHLNLTLSGNIDETDEAFTLWS 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
Q52172 PRELIMINARY; PRT; 90 AA.
AC Q52172;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE P2 PROTEIN (EC 1.14.13.7)
DE (PHENOL 2-MONOXYGENASE P2 COMPONENT).
GN PHHM.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-P35X / NCBI 9869;
RX MEDLINE; 95129877.
RA NG L.C., SHINGLER V., SZE C.C., POH C.L.;
RT "Cloning and sequences of the first eight genes of the chromosomally
RT encoded (methyl) phenol degradation pathway from Pseudomonas putida
RT P35X.";
RL Gene 151:29-36(1994).
CC -!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC DERIVATIVES.
CC -!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) +
CC H(2)O.
CC -!- COFACTOR: FAD FLAVOPROTEIN, AND REQUIRES FE(+2) FOR ACTIVITY (BY
CC SIMILARITY).
CC -!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED BY
CC P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
DR EMBL; X79063; CAA55662.1; -.
DR HSSP; P19731; IHQI.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
KW Flavoprotein; FAD; Iron.
SQ SEQUENCE 90 AA; 10519 MW; D9015E61 CRC32;

Query Match 54.8%; Score 247; DB 2; Length 90;

```

```

Best Local Similarity 52.3%; Pred. No. 1.4e-18;
Matches 46; Conservative 20; Mismatches 22; Indels 0; Gaps 0;

QY 2 SNVFIAFOANEDSRPIVDADVADNPRAVVVSPGMVKIDAPDLTIRRETEELTGTRFD 61
Db 3 SLVYIAFOQNDNARYLVEAIQDNPHAVVQHHPAMIRIEAEKRLREIRVEEENLGRAND 62

QY 62 LQOLQVNLITLSGHIDEDDEFTLSWSH 89
Db 63 VQEMLVDTVITIGNIDEDDDRFVLEWKN 90

RESULT 6
Q52163
ID Q52163 PRELIMINARY; PRT; 91 AA.
AC Q52163;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PHENOL HYDROXYLASE P2 PROTEIN (EC 1.14.13.7)
DE {PHENOL 2-MONOOXYGENASE P2 COMPONENT}.
DE PHLC.
GN Pseudomonas putida.
OS Pseudomonas putida.
OG Plasmid pGHL.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE; 95272534.
RA HERMANN H., MUELLER C., SCHMIDT I., MAHNKE J., PETRUSCHKA L.,
RA HAHNKE K.;
RT "Localization and organization of phenol degradation genes of
RT Pseudomonas putida strain H."
RL Mol. Gen. Genet. 247:240-246(1995).
CC -!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED
CC DERIVATIVES.
CC -!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) +
CC H(2)O.
CC -!- COFACTOR: FAD FLAVOPROTEIN, AND REQUIRES FE(+2) FOR ACTIVITY (BY
CC SIMILARITY).
CC -!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.
CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED BY
CC P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.
DR EMBL; X60765; CAA56742.1; -.
DR HSSP; P19731; 1HQI.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;
KW Flavoprotein; FAD; Iron; Plasmid.
SQ SEQUENCE 91 AA; 10661 MW; 129BLFDF CRC32;

Query Match 53.9%; Score 243; DB 2; Length 91;
Best Local Similarity 50.0%; Pred. No. 3.7e-18;
Matches 44; Conservative 21; Mismatches 23; Indels 0; Gaps 0;

QY 2 SNVFIAFOANEDSRPIVDADVADNPRAVVVSPGMVKIDAPDLTIRRETEELTGTRFD 61
Db 3 SLVYIAFOQNDNARYLVEAIQDNPHAVVQHHPAMIRIEAEKRLREIRVEEENLGRAND 62

QY 62 LQOLQVNLITLSGHIDEDDEFTLSWSH 89
Db 63 VQEMLVDTVITIGNIDEDDDRFVLEWKN 90

RESULT 7
Q32430
ID Q32430 PRELIMINARY; PRT; 89 AA.
AC Q32430;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE DMS OXYGENASE COMPONENT.
GN DSOC.

```

```

OS Acinetobacter sp.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Moraxellaceae; Acinetobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=20B;
RX MEDLINE; 98005684.
RA HORINOCHI M., KASUGA K., NOJIRI H., YAMANE H., OMORI T.;
RT "Cloning and characterization of genes encoding an enzyme which
RT oxidizes dimethyl sulfide in Acinetobacter sp. strain 20B."
RL FEMS Microbiol. Lett. 155:99-105(1997).
DR EMBL; D85083; BAA23332.1; -.
DR HSSP; P19731; 1HQI.
SQ SEQUENCE 89 AA; 10135 MW; A50444FF CRC32;

```

```

Query Match 51.7%; Score 233; DB 2; Length 89;
Best Local Similarity 46.0%; Pred. No. 3.9e-17;
Matches 40; Conservative 22; Mismatches 25; Indels 0; Gaps 0;

QY 2 SNVFIAFOANEDSRPIVDADVADNPRAVVVSPGMVKIDAPDLTIRRETEELTGTRFD 61
Db 3 SKVYLALQDNDTSRYITIEAEIQDNPQATIQYLPAMIRVESTGELVVRVRAETVSEKLGQNW 62

QY 62 LQOLQVNLITLSGHIDEDDEFTLSWS 88
Db 63 IQELQLNMITLGGNVEDDDSFILKWN 89

```

```

RESULT 8
Q43980
ID Q43980 PRELIMINARY; PRT; 89 AA.
AC Q43980;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
DE PHENOLHYDROXYLASE COMPONENT.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Moraxellaceae; Acinetobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB8250;
RX MEDLINE; 96154937.
RA EHRT S., SCHIRMER F., HILLEN W.;
RT "Genetic organization, nucleotide sequence and regulation of
RT dioxygenase in Acinetobacter calcoaceticus and catechol 1,2-
RL Mol. Microbiol. 18:13-20(1995).
DR EMBL; Z36909; CAA85382.1; -.
DR HSSP; P19731; 1HQI.
SQ SEQUENCE 89 AA; 10136 MW; AA99E933 CRC32;

```

```

Query Match 51.4%; Score 232; DB 2; Length 89;
Best Local Similarity 46.0%; Pred. No. 4.9e-17;
Matches 40; Conservative 21; Mismatches 26; Indels 0; Gaps 0;

QY 2 SNVFIAFOANEDSRPIVDADVADNPRAVVVSPGMVKIDAPDLTIRRETEELTGTRFD 61
Db 3 SKVYLALQDNDTSRYITIEAEIQDNPQATIQYLPAMIRVESTGELVVRVRAETVSEKLGQNW 62

QY 62 LQOLQVNLITLSGHIDEDDEFTLSWS 88
Db 63 IQELQLNMITLGGNVEDDDSFILKWN 89

```

```

RESULT 9
O87801
ID O87801 PRELIMINARY; PRT; 110 AA.
AC O87801;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)

```

DT	01-NOV-1996 (Tremblrel. 01, Created)
DT	01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT	01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE	TOLUENE-3-MONOOXYGENASE FERREDOXIN PROTEIN.
GN	TRUV.
OS	Burkholderia pickettii (Pseudomonas pickettii).
OC	Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC	Ralstonia.
RN	[1]
RN	SEQUENCE FROM N.A.
RC	STRAIN=PKO1;
RX	MEDLINE; 95172404.
RA	BYRNE A.M., KUKOR J.J., OLSEN R.H.;
RT	"Sequence analysis of the gene cluster encoding toluene-3-
RT	monooxygenase from Pseudomonas pickettii PKO1.";
RL	Gene 154:65-70(1995).
DR	EMBL; U04052; AAB09621.1; -.
KW	Monooxygenase.
SQ	SEQUENCE 104 AA; 11703 MW; 31CA0AC7 CRC32;
Query Match                      24.4%; Score 110; DB 2; Length 104;	
Best Local Similarity         32.0%; Pred. No. 0.00021;	
Matches         24; Conservative         20; Mismatches         29; Indels         2; Gaps         2;	
QY	9 QANEDSRPIVDIAVNPN--RAVVVESPGMYKIDAPDKLTIRRTIEELTGTFRDLOQLQ 66 :: : ::::     :::: ::  ::  ::      : :: :
Db	20 RASSTITAGVIEAAQEDNFGKSIRIDDKLAYVRIDTDGELLRRATLEALGRPFKMSELE 79
QY	67 VNLITLSGHIDEHDD 81     :      :
Db	80 VNLSFAGRIETD 94
RESULT 12	
O69181	PRELIMINARY; PRT; 105 AA.
ID O69181	
AC O69181;	
DT 01-AUG-1998 (Tremblrel. 07, Created)	
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)	
DT 01-AUG-1998 (Tremblrel. 07, Last annotation update)	
DE PUTATIVE HYDROXYLASE COMPONENT.	
GN PHLN.	
OS Alcaligenes eutrophus.	
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;	
OC Ralstonia.	
RN [1]	
RP SEQUENCE FROM N.A.	
RC STRAIN=JMP134;	
RA HARKER A.R., AYVOUBI P.J.;	
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.	
DR EMBL; AF065891; AAC77383.1; --	
SQ SEQUENCE 105 AA; 11913 MW; 3971DE3F CRC32;	
Query Match                      23.3%; Score 105; DB 2; Length 105;	
Best Local Similarity         32.0%; Pred. No. 0.00069;	
Matches         24; Conservative         17; Mismatches         32; Indels         2; Gaps         1;	
QY	9 QANEDSRPIVDIAVNPN--RAVVVESPGMYKIDAPDKLTIRRTIEELTGTFRDLOQLQ 66 :: : ::::     :::: ::  ::  ::      : :: :
Db	21 RASSTITAGVIEAAQEDNFGKEIRIVDDKLAYVRIDTDGELLRRATLEALGRPFKMSELE 80
QY	67 VNLITLSGHIDEHDD 81     :      :
Db	81 VNLSFAGRIETD 95
RESULT 13	
O92ET4	PRELIMINARY; PRT; 101 AA.
ID O92ET4	
AC O92ET4;	
DT 01-MAY-1999 (Tremblrel. 10, Created)	



```

DE 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE COUPLING/EFFECTOR PROTEIN.
GN XAMOD.
OS Xanthobacter sp. Py2.
OC Bacteria; Proteobacteria; alpha subdivision; Ancylobacter group;
OC Xanthobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PY2;
RA ZHOU N.Y., CHAN KWO CHION C.K., LEAK D.J.;
RT "The alkene monooxygenase from Xanthobacter Py2 is closely related to
RT aromatic monooxygenase and catalyses aromatic monohydroxylation of
RT benzene, toluene and phenol.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ012090; CAA09914.1; -.
SQ SEQUENCE 101 AA; 11194 MW; F4F59639 CRC32;

Query Match 20.0%; Score 90; DB 2; Length 101;
Best Local Similarity 24.2%; Pred. No. 0.023;
Matches 24; Conservative 19; Mismatches 42; Indels 14; Gaps 3;

QY 1 MSNVFI-----AFQANEDSRIVDAIVADNP--RAVVESPGMKIDAPDRLTIR 48
Db 1 MSNAIVDDMDENLVGPVIRAGDLADAVDAVDNPGKEVHIERGDYVRIHTDRDCRLT 60
QY 49 RTIEELGTRELDLQQLVNLTITLSGHIDEDEDEFTLSW 87
Db 61 RASIEQALGRSFVLAIAEAMSFKGRMSSDSE--MRW 97

RESULT 14
ID O07071 PRELIMINARY; PRT; 104 AA.
AC O07071;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE TBHD.
GN TBHD.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AA1;
RA MA Y., HERSON D.S.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF001356; AAB58743.1; -.
SQ SEQUENCE 104 AA; 11790 MW; 40251740 CRC32;

Query Match 18.6%; Score 84; DB 2; Length 104;
Best Local Similarity 34.3%; Pred. No. 0.098;
Matches 23; Conservative 10; Mismatches 28; Indels 6; Gaps 2;

QY 17 IVDAIV----ADNP--RAVVESPGMKIDAPDRLTIRRTIEELTGTFRDLQQLQVNL 70
Db 24 LVDAFVEARADNPGEICVDKRAYIRIDTGELLIPRETIERALGRFFKMPDLEVELS 83
QY 71 TLSGHID 77
Db 84 SPAGRIE 90

RESULT 15
QYX8M6
ID QYX8M6 PRELIMINARY; PRT; 862 AA.
AC QYX8M6;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

```

```

DE PUTATIVE LARGE ATP-BINDING PROTEIN.
GN SCE94.20.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA OLIVER K., HARRIS D.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA BENTLEY S.D., PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE; 97000351.
RA KINASHI H., HOPWOOD D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL049628; CAB40869.1; -.
KW ATP-binding.
SQ SEQUENCE 862 AA; 94537 MW; 631192DD CRC32;

Query Match 18.5%; Score 83.5; DB 2; Length 862;
Best Local Similarity 28.6%; Pred. No. 1.4;
Matches 18; Conservative 17; Mismatches 27; Indels 1; Gaps 1;

QY 27 RAVVESPGMKIDAPDRLTIRRTIEELTGTFRDLQQLQVNLITLSGHIDEDEDEFTLS 86
Db 556 RALLRS-GILREPARGRVDLHRTFDYLGARLAYQEMDFOLLVNHHLDEWDVILLA 614
QY 87 WSH 89
Db 615 LAH 617

Search completed: September 26, 2000, 20:28:54
Job time: 1013 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:30:04 ; Search time 58.11 Seconds  
(without alignments)  
275.174 Million cell updates/sec

Title: US-09-430-029-5

Perfect score: 2873

Sequence: 1 MDITPLKKLKLKLDRYAAMT.....FDGSEDQKNFAWRGQATRN 516

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 85661 seqs, 30989116 residues

Total number of hits satisfying chosen parameters: 85661

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_38.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1846.5	64.3	517	1	DMPN_PSESP
2	374.5	13.0	499	1	TMOA_PSEME
3	234.5	8.9	527	1	MEMA_METCA
4	211	7.3	525	1	TMOA_METTR
5	152	5.3	326	1	TMOE_PSEME
6	115	4.0	852	1	POL_BIVAU
7	107.5	3.7	852	1	POL_BIVJ
8	104	3.6	549	1	AMY_BACST
9	103	3.6	526	1	MLHL_ARATH
10	103	3.6	549	1	MCRA_METTH
11	99	3.4	549	1	MCRA_METTM
12	95	3.3	782	1	OSTA_HAEIN
13	94.5	3.3	1306	1	ACE_HUMAN
14	92.5	3.2	842	1	DMPN_LACDL
15	90.5	3.2	1256	1	KRP_STRSU
16	89.5	3.1	971	1	AGLU_ASPNG
17	89.5	3.1	985	1	ATSB_KLEAE
18	89	3.1	405	1	ATSB_KLEAE
19	89	3.1	670	1	TBUD_BURPL
20	89	3.1	875	1	VP34_YEAST
21	88	3.1	582	1	ASN2_PEA
22	88	3.1	713	1	DC12_ECOLI
23	87.5	3.0	702	1	ARYA_MANSE
24	87.5	3.0	1009	1	AMPM_HELVI
25	87	3.0	781	1	GCSL_CABEL
26	87	3.0	819	1	STL_CHLTR
27	86.5	3.0	666	1	Y032_MYCGE
28	86.5	3.0	815	1	GYRB_MYXXA
29	86	3.0	592	1	INV2_DAUCA
30	85.5	3.0	483	1	PHR_ANANT
31	84.5	2.9	585	1	ASN2_LOTJA
32	84.5	2.9	604	1	IAP1_HUMAN
33	84.5	2.9	723	1	SYM_PYRHO

34 84.5 2.9 1237 1 YDY2\_SCHPO  
35 84 2.9 535 1 YGIS\_ECOLI  
36 84 2.9 585 1 ASN1\_LOTJA  
37 84 2.9 619 1 PPKK\_HAECA  
38 84 2.9 621 1 HTPG\_RICPR  
39 84 2.9 757 1 DHET\_GLOSU  
40 83.5 2.9 605 1 MALZ\_ECOLI  
41 83 2.9 760 1 SPOT\_SYNK3  
42 82.5 2.9 557 1 G6PI\_KLULA  
43 82.5 2.9 585 1 YH67\_CABEL  
44 82.5 2.9 607 1 SYNC\_YEAST  
45 82.5 2.9 850 1 CYAA\_YERPE

## ALIGNMENTS

## RESULT 1

DMPN\_PSESP  
ID DMPN\_PSESP STANDARD; PRT; 517 AA.  
AC P19732;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-FEB-2000 (Rel. 39, Last annotation update)  
DE PHENOL HYDROXYLASE P3 PROTEIN (EC 1.14.13.7) (PHENOL 2-MONOOXYGENASE  
DE P3 COMPONENT).  
DE P3 COMPONENT).  
GN DMPN OR PHEM.  
OS Pseudomonas sp. (strain CF600).  
OG Plasmid pVil150.  
OC Bacteria; Proteobacteria.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 91072230.  
RA Nordlund I., Powlowski J., Shingler V.;  
RT "Complete nucleotide sequence and polypeptide analysis of  
RT multicomponent phenol hydroxylase from Pseudomonas sp. strain  
RT CF600".  
RL J. Bacteriol. 172:6826-6833(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BH;  
RA Takeo M., Maeda Y., Okada H., Miyama K., Mori K., Ike M.,  
RA Fujita M.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED  
CC DERIVATIVES. P3 IS REQUIRED FOR GROWTH ON PHENOL, AND FOR  
CC IN VITRO PHENOL HYDROXYLASE ACTIVITY.  
CC -!- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) + H(2)O.  
CC -!- COFACTOR: FAD, AND REQUIRES FE(+2) FOR ACTIVITY.  
CC -!- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.  
CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED  
CC BY P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M60276; AAA25942.1; -  
DR EMBL; D28864; BAA06017.1; -  
DR PIR; D37831; D37831.  
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;  
KW NADP; Flavoprotein; FAD; Iron; Plasmid.  
SQ SEQUENCE 517 AA; 60522 MW; 4EED4EB40ED73F9C CRC64;

Query Match 64.3%; Score 1846.5; DB 1; Length 517;  
Best Local Similarity 64.8%; Pred. No. 2.8e-145;  
Matches 328; Conservative 63; Mismatches 106; Indels 9; Gaps 1;



RA Rosenzweig A.C., Frederick C.A., Lippard S.J., Nordlund P.;  
RT "Crystal structure of a bacterial non-haem iron hydroxylase that  
RL catalyses the biological oxidation of methane."; Nature 366:537-543(1993).  
CC -!- FUNCTION: RESPONSIBLE FOR THE INITIAL OXYGENATION OF METHANE TO  
CC METHANOL IN METHANOTROPHS. IT ALSO CATALYZES THE MONOHYDROXYLATION  
CC OF A VARIETY OF UNACTIVATED ALKENES, ALICYCLIC, AROMATIC AND  
CC HETEROCYCLIC COMPOUNDS.  
CC -!- CATALYTIC ACTIVITY: METHANE + NAD(P)H + O(2) = METHANOL +  
CC NAD(P)(+) + H(2)O.  
CC -!- COFACTOR: BINDS TWO IRON ATOMS.  
CC -!- SUBUNIT: M.CAPSULATUS HAS TWO FORMS OF METHANE MONOOXYGENASE,  
CC A SOLUBLE AND A MEMBRANE-BOUND TYPE. THE SOLUBLE TYPE CONSISTS  
CC OF THREE COMPONENTS (A, B AND C): PROTEIN A, COMPRISING THREE  
CC CHAINS, IN AN ALPHA-2, BETA-2, GAMMA-2 CONFIGURATION, IS A  
CC NONHEME IRON PROTEIN CONTAINING AN UNUSUAL MU-HYDROXO BRIDGE  
CC STRUCTURE AT ITS ACTIVE SITE AND INTERACTS WITH BOTH OXYGEN AND  
CC METHANE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M58499; AAA25385.1; -;  
CC EMBL; M90050; AAB62392.1; -;  
CC PIR; JQ0702; JQ0702.  
CC -----  
CC Oxidoreductase; Monooxygenase; NADP; One-carbon metabolism; Iron.  
CC FT ACT\_SITE 151 151 POTENTIAL.  
CC FT METAL 114 114 IRON 1.  
CC FT METAL 144 144 IRON 1.  
CC FT METAL 147 147 IRON 1.  
CC FT METAL 209 209 IRON 2.  
CC FT METAL 243 243 IRON 2.  
CC FT METAL 246 246 IRON 2.  
CC SEQUENCE 527 AA; 60630 MW; 44DBB7A325423049 CRC64;  
CC -----  
CC Query Match 8.9%; Score 254.5; DB 1; Length 527;  
CC Best Local Similarity 24.2%; Pred. No. 1.3e-13;  
CC Matches 109; Conservative 65; Mismatches 179; Indels 97; Gaps 22;  
CC -----  
CC QY 54 DPFLRTMDAYWYQGEKELLYAVID-AFTONNAPFLGVSDARYINALKFLQGVTPLEYL 112  
CC Db EQFKLIAKEYARMEAVKDERQSGSLQVALTRLNAGVRV-HPKWNTEMKVV----- 106  
CC -----  
CC QY 113 AHRGFAHVGRH-----FTG---EGARIACQ-----MQSIDELEHYOTETHAMSTYKFF-- 158  
CC Db 107 --SNELEVGEYNATIAATGMLWDSQAQKNGYLAQVLDEIRH-----THQCAVYVYPAK 160  
CC -----  
CC QY 159 NG---FHHSNQWDFR---VWYLSVPKSEFFEDAYSSGFEFLTAVSFSE-----EYVLT 205  
CC Db 161 NGQDPAGHNDARRTRTIGPLW--KGMKRVFSDFGTSF-----DAVECSNLQLVGEACFT 213  
CC -----  
CC QY 206 NLLFPFMSGAAVNGMDSTVTTFGSAQSDSRHMTLGTECIKFLLEQDPNNVPIVQRWID 265  
CC Db 214 NPLIVAVTVERAANGDEITPTVFLSIEFDELHMHANGVQTV-VSTANDPASKAYLNTDLN 272  
CC -----  
CC QY 266 KWFNFGYKLLT-LVAMMDYMWOPKRVMSWRESWEMYA-EONGGALFKDLARYGIREPKGW 323  
CC Db 273 NAFWTQQKYFTVPLGMLFEYSGKFKVPEWPKTWNRWVYEDWGGIWLGRLGKYGVSFSL 332  
CC -----  
CC QY 324 QDACGCKDHISHQAWSTFYGFENAAAFHTVPTDEMWLSKAYPDSFDRYRPRFDHWG 383  
CC Db 333 KDAKQDAYWAHHDLVLLAYALMPTGFFRLALPDQEMEFWEANYPGWYD-HYKGIYEW- 390  
CC -----  
CC QY 384 EQARAGNRFPYKTLPMLCQTCQIP-----MLFTEPENGTKI-----GA- 421  
CC Db 391 -RARG-----CEDPSSGGFTPLMFWFIENNHPYIDRVSVQVFPFCPSLAKGAS 434

QY 422 --RESNYLGNKFHFCSDHCKDIFDHEPOKY 449  
Db 435 TLRVHEYNGQMHFTSPDOWGERWKLAPERY 464  
-----  
RESULT 4  
MEMA\_METTR STANDARD; PRT; 525 AA.  
AC P27353;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last annotation update)  
DE METHANE MONOOXYGENASE COMPONENT A ALPHA CHAIN (EC 1.14.13.25) (METHANE  
DE HYDROXYLASE).  
GN MMOX.  
OS Methylosinus trichosporium.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Methylocystaceae; Methylosinus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OB3b;  
RX MEDLINE; 91251762.  
RA Cardy D.L.N., Laidler V., Salmond G.P.C., Murrell J.C.;  
RT "Molecular analysis of the methane monooxygenase (MMO) gene cluster  
RT of Methylosinus trichosporium OB3b."; Mol. Microbiol. 5:335-342(1991).  
RL Mol. Microbiol. 5:335-342(1991).  
CC -!- FUNCTION: RESPONSIBLE FOR THE INITIAL OXYGENATION OF METHANE TO  
CC METHANOL IN METHANOTROPHS. IT ALSO CATALYZES THE MONOHYDROXYLATION  
CC OF A VARIETY OF UNACTIVATED ALKENES, ALICYCLIC, AROMATIC AND  
CC HETEROCYCLIC COMPOUNDS.  
CC -!- CATALYTIC ACTIVITY: METHANE + NAD(P)H + O(2) = METHANOL +  
CC NAD(P)(+) + H(2)O.  
CC -!- COFACTOR: BINDS TWO IRON ATOMS.  
CC -!- SUBUNIT: M.TRICHOSPORIUM HAS TWO FORMS OF METHANE MONOOXYGENASE,  
CC A SOLUBLE AND A MEMBRANE-BOUND TYPE. THE SOLUBLE TYPE CONSISTS  
CC OF THREE COMPONENTS (A, B AND C): PROTEIN A, COMPRISING THREE  
CC CHAINS, IN AN ALPHA-2, BETA-2, GAMMA-2 CONFIGURATION, IS A  
CC NONHEME IRON PROTEIN CONTAINING AN UNUSUAL MU-HYDROXO BRIDGE  
CC STRUCTURE AT ITS ACTIVE SITE AND INTERACTS WITH BOTH OXYGEN AND  
CC METHANE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X55394; CAA39068.1; -;  
CC PIR; S15207; S15207.  
CC -----  
CC KW Oxidoreductase; Monooxygenase; NADP; One-carbon metabolism; Iron.  
CC FT ACT\_SITE 151 151 POTENTIAL.  
CC FT METAL 114 114 IRON 1 (BY SIMILARITY).  
CC FT METAL 144 144 IRON 1 (BY SIMILARITY).  
CC FT METAL 147 147 IRON 1 (BY SIMILARITY).  
CC FT METAL 209 209 IRON 2 (BY SIMILARITY).  
CC FT METAL 243 243 IRON 2 (BY SIMILARITY).  
CC FT METAL 246 246 IRON 2 (BY SIMILARITY).  
CC SEQUENCE 525 AA; 59989 MW; FD5227DEEFAF9DDA CRC64;  
CC -----  
CC Query Match 7.3%; Score 211; DB 1; Length 525;  
CC Best Local Similarity 21.0%; Pred. No. 5.2e-10;  
CC Matches 96; Conservative 71; Mismatches 196; Indels 94; Gaps 19;  
CC -----  
CC QY 45 KTHMDKWDPFLRTMDAYWYQGEKELLY-AVIDAFTQNNNA----- 86  
CC Db 49 KYHMANETKEQFKVIAKEYARMEAAKDERQFGTLLDGLTRLGAGNKVHPRWGETMKVLSN 108  
CC -----  
CC QY 87 FLGVSDARYINALKFLQGVTPLE-----YLAHRGFAHVGRHFTGEGARIACQMSIDEIR 142  
CC Db 109 FLEVGEYNAIAASAMLDWSATAAEQKNGYLA-----QVLDEIR 146

```
QY 143 HYQETHAMSTYKPFNGFHH-----NOWFDRWYLVKPSFFEDAY-SSGPF 192
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
Db 147 H-THQCAFINHYSKHVHDPAHNDARTRAIGPLW--KGMKRVFADGFISRDV 200
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 193 LTAYSPFSEYVLTNLPVPMGSAAYNGDMSTVTFGSAQDSRHMVLGTEICF 252
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
Db 201 SVNQLQVGDTCFTNPLIVAVTEWAIGNGDEITPTVFTSVETDELHMHANG 259
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 253 DPDNVPIVQWIDKFWGRGRLT-LVAMMDYMQPKRVMSWRESWEMAYBQN-G 310
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
Db 260 DPASAKFLNTDLNNAFTQKYFTPVLYLEFGSKFVFPWVATNKNWSEDMG 319
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 311 DLARYGIREPKGWDAGEGDKHISHQAWSTFYGFNAASAFHTW-----V 362
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
Db 320 RLKYGIV-ESKVLROAKRDYWAHDL-----ALAAAYAMWPLAFARLAPD 370
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 363 LSAXYPSDFRYPRFDHWG-----EGARAGNRYMYKTL-----MLC 413
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
Db 371 FEANYPGWAD-HYKGFENWKGLGYEDPKSGFIPYQWILLANGHDVYIDRV 426
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 414 GNPTKIGA-RESNYLGNKFECHSCDKDIFDHEPOKY 449
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
Db 427 SLAKGTGLRVHEFGKKHSLTDWGERQWLIPEPY 463
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 5
TMOE_PSEME STANDARD; PRT; 326 AA.
AC Q00460;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TOLUENE-4-MONOOXYGENASE SYSTEM PROTEIN E (EC 1.14.13.-).
GN TMOE.
OS Pseudomonas mendocina.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
OC Pseudomonas.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-14.
RC STRAIN-KF1.
RX MEDLINE; 91358306.
RA Yen K.-M., Karl M.R., Blatt L.M., Simon M.J., Winter R.B.,
RA Fausset P.R., Lu H.S., Harcourt A.A., Chen K.K.;
RT "Cloning and characterization of a Pseudomonas mendocina Kx1 gene
RT cluster encoding toluene-4-monooxygenase.";
RL J. Bacteriol. 173:5315-5327(1991).
CC -1- FUNCTION: HYDROXYLATES TOLUENE TO FORM P-CRESOL.
CC -1- COFACTOR: FAD; REQUIRES FE(2) FOR ACTIVITY.
CC -1- PATHWAY: FIRST STEP IN TOLUENE DEGRADATION.
CC -1- SUBUNIT: THE MULTICOMPONENT ENZYME TOLUENE-4-MONOOXYGENASE
CC IS FORMED BY THE TMOA, TMOB, TMOG, TMOH, TMOI AND TMOJ
CC POLYPEPTIDES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL; M65106; AAA26003.1; -
DR Aromatic hydrocarbons catabolism; Oxidoreductase; Flavoprotein;
KW Monooxygenase; FAD; Iron.
FT INIT_MET 0
SQ SEQUENCE 326 AA; 38255 MW; 084EE106D06FB5EC CRC64;
```

Query Match 5.3%; Score 152; DB 1; Length 326;  
Best Local Similarity 24.5%; Pred. No. 2.1e-05;  
Matches 64; Conservative 34; Mismatches 131; Indels 32; Gaps 10;

```
QY 25 WE-TTYQMDKVF-PYDRIEGIKIHDKWDKWDPPFLTMDAYWKYQGEKEKKLYAVIDAFT 82
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
Db 42 WELSPDSMNLYKQYRNASPLKHDNWDFAFDPOQLVVTYTNLMQDGQESVQSLEFDQFN 101
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 83 ONNAFLGVSDARYINALKFLQGVTPLEYLAH--RGFAHVGVRHFTGGARIACOMQSID 139
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
Db 102 EREHDMQVREGWEHTMARC-----SPLRYLFHCLOMSAAYVOOMAPASTISNCC 157
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 140 ELR---HYQTTHAMS-TYNKFENGFSHNSQFDRVWYLVKPSFFEDAYSSGPFELTA 195
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
Db 158 SLRWLTHATYTHSLTYPDAGLGEHERELWEKEPQGGL-RELME-----KQLTA 208
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 196 VSFSEFYVLTNL-----LVEFFMSGAAAYNGDMSTVTFGSAQDSRHMVLGIECI 246
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
Db 209 FDWGEAFVSLNVKPMVIESIFKPLQQAWEENNNTLLPLLIDSQLKDAERHSRWSKALV 268
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
QY 247 KFLLEQDDPNVPIVQWIDKW 267
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
Db 269 KHALE-NPDNHAIVIEGWIEKW 288
| | | | | : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 6
POL_BLVAU STANDARD; PRT; 852 AA.
AC P25059;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE].
GN POL.
OS Bovine leukemia virus (Australian isolate) (BLV).
OC Viruses; Retroviruses; Retroviridae; BLV-HIVL retroviruses.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90362060.
RA Coulston J., Naif H., Brandon R., Kumar S., Khan S., Daniel R.C.W.,
RA Lavin M.F.;
RT "Molecular cloning and sequencing of an Australian isolate of
RT proviral bovine leukaemia virus DNA: comparison with other
RT isolates.";
RL J. Gen. Virol. 71:1737-1746(1990).
CC -1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: VERY STRONG, WITH BOVINE LEUKEMIA VIRUS ENV
CC POLYPROTEIN FROM OTHER ISOLATES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC
CC EMBL; D00647; BAA00544.1; -
DR PIR; JQ0555; GNJGGA.
DR HSP; P04585; 1RT1.
DR PFAM; PF00552; integrase; 1.
DR PFAM; PF00075; rnaaseH; 1.
DR PFAM; PF00665; rve; 1.
DR PFAM; PF00078; rvt; 1.
DR PFAM; PF00078; rvt; 1.
KW Transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;
KW Endonuclease; Polyprotein.
SQ SEQUENCE 852 AA; 95441 MW; CD69FE59E1208A08 CRC64;
```

Query Match 4.0%; Score 115; DB 1; Length 852;  
Best Local Similarity 18.7%; Pred. No. 0.085;  
Matches 134; Conservative 81; Mismatches 252; Indels 248; Gaps 36;

```
QY 3 TPTLKKLGLKDRYAAMTRGLGWET-----TYQPM-----DKVFPDYREGIKIH 47
| | | | | : : : | | : : : : : : : : : : : : : : : : : : : : :
Db 169 SPTEERLQCYQTMAHLRLDGLFQVASEKTRQTFSPVFLGMVHERMTVYQSLPTLQIS 228
| | | | | : : : | | : : : : : : : : : : : : : : : : : : : : :
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02120; AAA42785.1; -.
DR PIR; A03960; GNLJGB.
DR HSSP; P04585; IRT1.
DR PFAM; PF00552; integrase; 1.
DR PFAM; PF00075; rnaaseh; 1.
DR PFAM; PF00665; rve; 1.
DR PFAM; PF00078; rvt; 1.
KW transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;
KW Endonuclease; Polyprotein.
SQ SEQUENCE 852 AA; 95182 MW; A7A63524A18C458D CRC64;

Query Match      3.7%; Score 107.5; DB 1; Length 852;
Best Local Similarity 19.0%; Pred. No. 0.35;
Matches 115; Conservative 70; Mismatches 214; Indels 207; Gaps 30;

QY 3 TPTLKKLGLKDRYAMTRGLGNET-----TYQPM-----DKVFPYDRYEIGIKH 47
Db 169 SPTEFORSCYQALAAARLDLGFQVASEKTSQTPSPVFLGQMVHEQIVTYOSLPTQIS 228

QY 48 DWDKWDVDFRLTM-----DAYWKYQGEKEK-----LYAVIDAFTONNAFLGVSDARV 96
Db 229 -----SPISLHQLQAVLGLDQWVSRGTPTRRPIQLLYSSLKRRHDPRAITQLSPEQ-- 280

QY 97 NALKFLQGVTPLEYLHRG-----FAHVGHRFTGEGARTAC-OMGSIDELRHY 144
Db 281 -----LQGIARELQALSHNARSRYNEQEPFLAYV--HLTRAGSTLVLFQKGAQFPFLAYF 332

QY 145 QT-----ETHAMSTYNKFFNGFPHS-----NOWED----- 169
Db 333 QTPPLDQNASPWGLILLGLCCYLQALSSYAKPIILKYHNLPKTSLDNWIOSSEDPVQ 392

QY 170 -----RWVYLSVPKSFEDAYSSGFPE-----FLTAVSFSEYVLTNLLFVPMGSAAY 218
Db 393 ELLQLW-----POISSGQIOPPGPWKTLITRAEVFLTP-QFSPDPIPAALCL--FSDGATG 445

QY 219 NG-----DMSTVTFGSAQSDSHMTLGTICIKFLQDQDPDNPVIVQWID-KW 267
Db 446 RGAYCLWKDHLDFQAVPAPESAQKGLAGLLAG-----LAAAPPEPVNI---WVDSKY 496

QY 268 FWRGYRLTLVAMMDYMQPKVMSWR----- 294
Db 497 LYSLLRTLVLGA-----WLQPDVPVSYALLYKSLLRHPAIVGHRVSHSSASHPIASLNNY 552

QY 295 -----ESWEMYAEQNGALFK-----DLARYGIREPKGWQDACE-----GKDH 333
Db 553 VDQLPLETPEQWHKLTCHNSRALSRLSWPNPRISANDPRSPATLCETCQKLNITGGGKMT 612

QY 334 SHQAWS-----TFYGFNAAS-AFHTVPTDEMGWLSAKYPSDFRYRPRFD--- 380
Db 613 IQRGWAPNHIWQADITHYKYQFTYALHVFVDYSGATHASAKRGLTTQTITIEGLEAIV 672

QY 381 HWGEQAR-----AGNRFYMKTLPMLCQTCQIPMLFTEPGNPTKIGARES-----NYLGNKF 431
Db 673 HLGREPKLNTDQGANYSKIFVRFCQFGVSLSHHVPYNPNTSSGLDRTNGLLKLLSKY 732

QY 432 HFCSDH 437
Db 733 HLDEPH 738

RESULT 8
AMY_BACST STANDARD; PRT; 549 AA.
AC P06279; Q45519;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)

QY 48 DWDKWDVDFRLTM-----DAYWKYQGEKEK-----LYAVIDAFTONNAFLGVSDARY 96
Db 229 -----SPISLHQLQAVLGLDQWVSRGTPTRRPIQLLYSSLKRRHDPRAITQLSPEQ-- 281

QY 97 NALKFLQGVTPLEYLHRG-----FAHVGHRFTGEGARTAC-OMGSIDELRHY 144
Db 282 -----QGIARELQALSHNARSRYNEQEPFLAYV--HLTRAGSTLVLFQKGAQFPFLAYF 332

QY 145 QT-----ETHAMSTYNKFFNGFPHS-----NOWED----- 169
Db 333 QTPPLDQNASPWGLILLGLCCYLQALSSYAKTILKYHNLPKTSLDNWIOSSEDPVQ 392

QY 170 -----RWVYLSVPKSFEDAYSSGFPE-----FLTAVSFSEYVLTNLLFVPMGSAAY 218
Db 393 ELLQLW-----POISSGQIOPPGPWKTLITRAEVFLTP-QFSPDPIPAALCL--FSDGAAR 445

QY 219 NG-----DMSTVTFGSAQSDSHMTLGTICIKFLQDQDPDNPVIVQWID-KW 267
Db 446 RGAYCLWKDHLDFQAVPAPESAQKGLAGLLAGLAA---APPEPLMI-----WVDSKY 496

QY 268 FWRGYRLTLVAMMDYMQPKVMSWR----- 294
Db 497 LYSLLRTLVLGA-----WLQPDVPVSYALLYKSLLRHPAIVGHRVSHSSASHPIASLNNY 552

QY 295 -----ESWEMYAEQNGALFK-----DLARYGIREPKGWQDACE-----GKDH 333
Db 553 VDQLPLETPEQWHKLTCHNSRALSRLSWPNPRISANDPRSPATLCETCQKLNITGGGKMT 612

QY 334 SHQAWS-----TFYGFNAAS-AFHTVPTDEMGWLSAKYPSDFRYRPRFD--- 380
Db 613 IQRGWAPNHIWQADITHYKYQFTYALHVFVDYSGATHASAKRGLTTQTITIEGLEAIV 672

QY 381 HWGEQAR-----AGNRFYMKTLPMLCQTCQIPMLFTEPGNPTKIGARES-----NYLGNKF 431
Db 673 HLGREPKLNTDQGANYSKIFVRFCQFGVSLSHHVPNPTSSGLVERTNGLLKLLSKY 732

QY 432 HFCSDH-----CKDIFHEPQKYV-----QAW-----LPVHQIHQGNCFPPDAD----- 470
Db 733 HLDEPHLPMQALSRAALTHQINLPLKTRWELHSPPLAVISEGGTPKGSKLFLY 792

QY 471 --PGNEG--TDPAAVLIDYAVTMGRDN-----LDFGSEDOKN 505
Db 793 KLPGONRRRWGLPPLPAIVEASGALLATNPVVPWRLLKAFKCLKNDGPDAPN 847

POL_BLVJ
ID POL_BLVJ STANDARD; PRT; 852 AA.
AC P03361;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE (EC 2.7.7.49);
DE ENDONUCLEASE].
GN POL.
OS Bovine leukemia virus (Japanese isolate BLV-1) (BLV).
OC Viruses; Retroviral viruses; Retroviridae; BLV-HIV retroviruses.
KN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85140159.
RA Sagata N., Yasunaga T., Tsuzuku-Kawamura J., Ohishi K., Ogawa Y.,
RA Ikawa Y.;
RT "Complete nucleotide sequence of the genome of bovine leukemia virus;
RT its evolutionary relationship to other retroviruses.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:677-681 (1985).
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: VERY STRONG, WITH BOVINE LEUKEMIA VIRUS ENV
CC POLYPROTEIN FROM OTHER ISOLATES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02120; AAA42785.1; -.
DR PIR; A03960; GNLJGB.
DR HSSP; P04585; IRT1.
DR PFAM; PF00552; integrase; 1.
DR PFAM; PF00075; rnaaseh; 1.
DR PFAM; PF00665; rve; 1.
DR PFAM; PF00078; rvt; 1.
KW transferase; RNA-directed DNA polymerase; Hydrolase; Nuclease;
KW Endonuclease; Polyprotein.
SQ SEQUENCE 852 AA; 95182 MW; A7A63524A18C458D CRC64;

Query Match      3.7%; Score 107.5; DB 1; Length 852;
Best Local Similarity 19.0%; Pred. No. 0.35;
Matches 115; Conservative 70; Mismatches 214; Indels 207; Gaps 30;

QY 3 TPTLKKLGLKDRYAMTRGLGNET-----TYQPM-----DKVFPYDRYEIGIKH 47
Db 169 SPTEFORSCYQALAAARLDLGFQVASEKTSQTPSPVFLGQMVHEQIVTYOSLPTQIS 228

QY 48 DWDKWDVDFRLTM-----DAYWKYQGEKEK-----LYAVIDAFTONNAFLGVSDARV 96
Db 229 -----SPISLHQLQAVLGLDQWVSRGTPTRRPIQLLYSSLKRRHDPRAITQLSPEQ-- 280

QY 97 NALKFLQGVTPLEYLHRG-----FAHVGHRFTGEGARTAC-OMGSIDELRHY 144
Db 281 -----LQGIARELQALSHNARSRYNEQEPFLAYV--HLTRAGSTLVLFQKGAQFPFLAYF 332

QY 145 QT-----ETHAMSTYNKFFNGFPHS-----NOWED----- 169
Db 333 QTPPLDQNASPWGLILLGLCCYLQALSSYAKPIILKYHNLPKTSLDNWIOSSEDPVQ 392

QY 170 -----RWVYLSVPKSFEDAYSSGFPE-----FLTAVSFSEYVLTNLLFVPMGSAAY 218
Db 393 ELLQLW-----POISSGQIOPPGPWKTLITRAEVFLTP-QFSPDPIPAALCL--FSDGATG 445

QY 219 NG-----DMSTVTFGSAQSDSHMTLGTICIKFLQDQDPDNPVIVQWID-KW 267
Db 446 RGAYCLWKDHLDFQAVPAPESAQKGLAGLLAG-----LAAAPPEPVNI---WVDSKY 496

QY 268 FWRGYRLTLVAMMDYMQPKVMSWR----- 294
Db 497 LYSLLRTLVLGA-----WLQPDVPVSYALLYKSLLRHPAIVGHRVSHSSASHPIASLNNY 552

QY 295 -----ESWEMYAEQNGALFK-----DLARYGIREPKGWQDACE-----GKDH 333
Db 553 VDQLPLETPEQWHKLTCHNSRALSRLSWPNPRISANDPRSPATLCETCQKLNITGGGKMT 612

QY 334 SHQAWS-----TFYGFNAAS-AFHTVPTDEMGWLSAKYPSDFRYRPRFD--- 380
Db 613 IQRGWAPNHIWQADITHYKYQFTYALHVFVDYSGATHASAKRGLTTQTITIEGLEAIV 672

QY 381 HWGEQAR-----AGNRFYMKTLPMLCQTCQIPMLFTEPGNPTKIGARES-----NYLGNKF 431
Db 673 HLGREPKLNTDQGANYSKIFVRFCQFGVSLSHHVPYNPNTSSGLDRTNGLLKLLSKY 732

QY 432 HFCSDH 437
Db 733 HLDEPH 738

RESULT 8
AMY_BACST STANDARD; PRT; 549 AA.
AC P06279; Q45519;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
```

DE ALPHA-AMYLASE PRECURSOR (BC 3.2.1.1.1) (1,4-ALPHA-D-GLUCAN  
DN GLUCANOHYDROLASE).  
GN AMYS.  
OS Bacillus stearothermophilus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-39.  
RX MEDLINE; 85234394.  
RA Nakajima R., Imanaka T., Aiba S.;  
RT "Nucleotide sequence of the Bacillus stearothermophilus alpha-amylase  
RT gene.";  
RL J. Bacteriol. 163:401-406(1985).  
RN [2]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=DY5/PHI300;  
RX MEDLINE; 86008166.  
RA Ihara H., Sasaki T., Tsuboi A., Yamagata H., Tsukagoshi N., Uda S.;  
RT "Complete nucleotide sequence of a thermophilic alpha-amylase gene:  
RT homology between prokaryotic and eukaryotic alpha-amylases at the  
RT active sites.";  
RL J. Biochem. 98:95-103(1985).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NZ-3;  
RX MEDLINE; 86195857.  
RA Gray G.L., Mainzer S.E., Rey M.W., Lamsa M.H., Kindie K.L.,  
RA Carmona C., Requadt C.;  
RT "Structural genes encoding the thermophilic alpha-amylases of  
RT Bacillus stearothermophilus and Bacillus licheniformis.";  
RL J. Bacteriol. 166:635-643(1986).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Suominen I., Karp M., Lautamo J., Knowles J., Mantsaerlae P.;  
RT "Thermotable alpha amylase of Bacillus stearothermophilus: cloning,  
RT expression, and secretion by Escherichia coli.";  
RL (In) Chaloupka J., Krumphanz V. (eds.);  
RL Extracellular enzymes of microorganisms, pp.129-137, Plenum Press,  
RL New York (1987).  
RN [5]  
RP SEQUENCE OF 1-122 FROM N.A., AND SEQUENCE OF 35-48.  
RC STRAIN=DY-5;  
RX MEDLINE; 86059211.  
RA Tsukagoshi N., Iritani S., Sasaki T., Takemura T., Ihara H.,  
RA Idota Y., Yamagata H., Uda S.;  
RT "Efficient synthesis and secretion of a thermophilic alpha-amylase by  
RT protein-producing Bacillus brevis 47 carrying the Bacillus  
RT stearothermophilus amylase gene.";  
RL J. Bacteriol. 164:1182-1187(1985).  
CC [1]- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC  
CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.  
CC [1]- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO  
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M11450; AAA22235.1; -;  
DR EMBL; X02769; CAA26547.1; -;  
DR EMBL; M57457; AAA22227.1; -;  
DR EMBL; M13255; AAA22241.1; -;  
DR PIR; A00845; ALHSF.  
DR PIR; A24549; A24549.  
DR HSP; P06278; IBPL.  
DR PRAM; PF00128; alpha-amylase; 1.  
KW Hydrolase; Glycosidase; Carbohydrate metabolism; Signal.  
FT SIGNAL 1 34  
FT CHAIN 35 549 ALPHA-AMYLASE.

FT	ACT_SITE	268	268	BY SIMILARITY.
FT	ACT_SITE	272	272	BY SIMILARITY.
FT	ACT_SITE	365	365	BY SIMILARITY.
FT	CONFLICT	13	13	M -> V (IN REF. 3).
FT	CONFLICT	19	19	L -> W (IN REF. 3).
FT	CONFLICT	23	23	L -> S (IN REF. 2 AND 3).
FT	CONFLICT	31	31	P -> H (IN REF. 2 AND 5).
FT	CONFLICT	107	107	A -> T (IN REF. 2 AND 3).
FT	CONFLICT	167	167	T -> I (IN REF. 4).
FT	CONFLICT	179	179	P -> N (IN REF. 3).
FT	CONFLICT	251	251	S -> N (IN REF. 2, 3 AND 4).
FT	CONFLICT	260	260	TNI -> RTL (IN REF. 4).
FT	CONFLICT	284	284	D -> Y (IN REF. 2, 3 AND 4).
FT	CONFLICT	312	312	M -> T (IN REF. 2 AND 3).
FT	CONFLICT	338	338	T -> A (IN REF. 2 AND 3).
FT	CONFLICT	342	342	R -> S (IN REF. 3).
FT	CONFLICT	346	346	T -> N (IN REF. 3).
FT	CONFLICT	376	376	V -> C (IN REF. 2 AND 3).
FT	CONFLICT	526	527	WS -> RP (IN REF. 2).
FT	CONFLICT	527	527	S -> P (IN REF. 2).
FT	CONFLICT	535	535	D -> G (IN REF. 2 AND 3).
SO	SEQUENCE	549 AA;	62670 MW;	3A2DD93A955E79D3 CRC64;

Query Match 3.6%; Score 104; DB 1; Length 549;  
Best Local Similarity 18.3%; Pred. No. 0.39;  
Matches 82; Conservative 43; Mismatches 125; Indels 198; Gaps 24;

Qy	21	RGLG-WETTYQPMKVPFYDREYGI-----KIHDWDKW-----VDPFRITMDAY 63
Db	213	RGIGKAWDEVDTEGNYDLYMADLDHPEVVTLSKSGKGYVNTTNDGFRDLDAVKH 272
Qy	64	WKYQGEKEKKLYAVIDAFTQNNAFI--CVSDARYINALKFLQGVTPLEYLAHRCFAHVG 121
Db	273	IKF-----SFFPDWLSVRSQTKPLFTVG-----EYWSY----- 302
Qy	122	RHFTGEGARIACQMOSIDELRHYQETHEAMSTYKFFGEHHSNOWEDRVWVLSVPKSEFF 181
Db	303	-----DINKLHNYIMKTN--GTMSLFDAPLEN-------KFY 330
Qy	182	EDAYSSGPFELTAVSFSEYVLTNLLEVPFMSGAAAYNGDMSTVTTFGSAOSDSRHMTL 241
Db	331	TASKSGGTFDMRT-----LMTNTLM-----KQDPTL----- 356
Qy	242	GIECKFLEODPNVPVQWIDKWF-----WRGYRL----- 275
Db	357	---AVTFVDNHDTEPGQALQSWDPFKPLAYAFILTRQEGYPCVFGYGYGIPQYNIPS 413
Qy	276	-----TLVA-----MMMDYMPKRVMSW-RE-----SWEM 299
Db	414	LKSKIDPLLIARRDYAYGTQHDYLDHSDIIGTWREGVTEKPGSGLAALITDGGSKWMY 473
Qy	300	YAEONGGALFKDLARYGIREPKGWODACEGKDHSQWSTFYGPNAAAFHWVWV---T 356
Db	474	VGKHAGKVFVDLT--GNR-----SDTIVT-----INSDGWGEF-KVNGGSV-SVWVPRKTT 520
Qy	357	EDEMGWLSAKYP--DSEDRYRPRPDHW 382
Db	521	VSTTAWSTTRPTWTEDEFVRWTEPRLVAV 548

RESULT 9  
MLH1 ARATH  
ID MLH1 ARATH STANDARD; PRT; 526 AA.  
AC O49621; O22766;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE MLC PROTEIN HOMOLOG 1 (ATMLO-H1).  
GN MLC-H1 OR T10p11.12.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;



```
OC Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA Panstruga R.:
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-CV. COLUMBIA;
RC Kaplan N., Johnson D., Schutz K., Gnoj L., Hoffman J., Till S.,
RA de la Bastide M., Granat S., Hameed A., Gottesman T., Hasegawa A.,
RA Shohdy N., Parnell L., Dedhia N., Johnson A.F., Lodhi M.,
RA Martienssen R., Chen E.Y., Wilson R., McCombie W.R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE MLO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z95352; CAB08605.1; -.
DR EMBL; AC002330; AAC78258.1; -.
KW Transmembrane.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 288 308 POTENTIAL.
FT TRANSMEM 309 329 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 413 433 POTENTIAL.
FT TRANSMEM 61 61 E -> GVIVHTHQQ (IN REF. 2).
FT CONFLICT 206 339 MISSING (IN REF. 2).
FT CONFLICT 393 435 VTGFDSGNGVRYIVPRIVGFIQVLCYSYSLPLIYIV
FT (IN REF. 2).
FT SQ -> GLHSSALQLQYTASLRHRTGNKLIFTKKLENIL
FT (IN REF. 2).
FT SEQUENCE 526 AA; 59126 MW; 8E63C276A71A9768 CRC64;
Query Match 3.6%; Score 103; DB 1; Length 526;
Best Local Similarity 22.0%; Pred. No. 0.45;
Matches 71; Conservative 43; Mismatches 130; Indels 78; Gaps 17;
QY 45 KIHDMKWDVDFRLTMDAYKYQGE---KEKKLYAVIDAFTQNNAFNLGVSDARYI----- 96
Db 184 RIHQWKWEDSI-----ADEKFDPTALRKRRVTHVHNHAFIKEH-FUGIGKDSVILGWT 237
QY 97 -NALKLFLQGVTPLEYLAHR-GFAHVGRHFTGEGARIACQMOSIDELRHYQTETHAMSTY 154
Db 238 QSFLKQFVDSYTKSDYVTLRIGF--IMTHCKG-NPKLNFHKYMRALDDDFKQVVGISWY 294
QY 155 NKFF-----NGPHHSNWFDRVWLVSPKSFEDDAYSGPPEFLTAVSFSEYVLT 205
Db 295 LWIEVWIFLLINVGWH-----TYFWLAFI--PFALLAVGTLKLEHVIA 336
QY 206 NLLVFPMSGAYNGDM-----STVTFGSAQSDSRHMTLGIETKFTLLEQDPNPVIVQ 261
Db 337 QLAHEVAEKHVAIGDLVVKPSDEHFWFSKPO-----IVLYLHFILFQNAFEIAFF- 388
QY 262 RWIDKWFWRGVRLLTLVAMMDYMOPK-----RVMSWSRESWEMYA--EONGG-----AL 308
Db 389 ----FWIWTVTGFSICMGQVRYIVPRLVIGVFIVQLCSYSLPLIYIVSQMGSSFKKAI 444
QY 309 FKDLARVIREPKGWDACEK 330
Db 445 FEENVQVGL---VGWAQVKQK 463
RESULT 10
MCRA_METH
```

```
ID AC MCRA_METH STANDARD; PRT; 549 AA.
DT 027232; 050493;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE METHYL-COENZYME M REDUCTASE I ALPHA SUBUNIT (EC 1.8.-.-) (MCR I
DE ALPHA).
GN MCRA OR MTH11364.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE; 95014084.
RA Pihl T.D., Sharma S., Reeve J.N.;
RT *Growth phase-dependent transcription of the genes that encode the
RT two methyl coenzyme M reductase isoenzymes and N5-
RT methyltetrahydromethanopterin:coenzyme M methyltransferase in
RT Methanobacterium thermoautotrophicum delta H.*;
RL J. Bacteriol. 176:6384-6391(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE; 98037514.
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT *Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.*;
RL J. Bacteriol. 179:7135-7155(1997).
RN [3]
RP SEQUENCE OF 1-18.
RC STRAIN-DELTA H;
RX MEDLINE; 91099370.
RA Rospert S., Linder D., Ellermann J., Thauer R.K.;
RT *Two genetically distinct methyl-coenzyme M reductases in
RT Methanobacterium thermoautotrophicum strain Marburg and delta H.*;
RL Eur. J. Biochem. 194:871-877(1990).
CC -1- FUNCTION: THIS ENZYME COMPLEX CATALYZES THE FINAL STEP IN
CC METHANOGENESIS, WHICH IS THE TERMINAL STEP OF ANAEROBIC
CC DEGRADATION OF BIOMASS.
CC -1- CATALYTIC ACTIVITY: REDUCTION OF METHYL-COENZYME M (2-(METHYLTHIO)
CC ETHANESULFONIC ACID) WITH 7-MERCAPTOHEPTANOYLTHREONINE PHOSPHATE
CC TO METHANE AND AN HETERODISULFIDE (CH(3)-S-COM + H-S-HTP = CH(4) +
CC COM-S-S-HTP).
CC -1- COFACTOR: THE ENZYME COMPLEX BINDS TIGHTLY (BUT NOT COVALENTLY)
CC TO TWO MOLECULES OF COENZYME F430. F430 IS A YELLOW NICKEL
CC PORPHINOID.
CC -1- SUBUNIT: HEXAMER OF TWO ALPHA, TWO BETA, AND TWO GAMMA CHAINS.
CC -1- DEVELOPMENTAL STAGE: THERE ARE TWO MCR COMPLEXES IN THIS BACTERIA.
CC MCR II IS EXPRESSED IN THE EARLY GROWTH PHASE. LATE GROWTH CELLS
CC CONTAINS MOSTLY MCR I.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U10036; AAA73445.1; -.
DR EMBL; AE000885; AAB85653.1; -.
KW Methanogenesis; Oxidoreductase; Methylation; Multigene family.
FT INIT_MET 0
FT MOD_RES 256 256 METHYLATION (BY SIMILARITY).
FT MOD_RES 269 269 METHYLATION (BY SIMILARITY).
FT CONFLICT 287 287 D -> E (IN REF. 1).
```

Science 278:1457-1462(1997).

-i- FUNCTION: THIS ENZYME COMPLEX CATALYZES THE FINAL STEP IN METHANOGENESIS, WHICH IS THE TERMINAL STEP OF ANAEROBIC DEGRADATION OF BIOMASS.

-i- CATALYTIC ACTIVITY: REDUCTION OF METHYL-COENZYME M (2-(METHYLTHIO) ETHANESULFONIC ACID) WITH 7-MERCAPTOHEPTANOYLTHREONINE PHOSPHATE TO METHANE AND AN HETERODISULFIDE (CH(3)-S-COM + H-S-HTP = CH(4) + COM-S-S-HTP).

-i- COFACTOR: THE ENZYME COMPLEX BINDS TIGHTLY (BUT NOT COVALENTLY) TO TWO MOLECULES OF COENZYME F430. F430 IS A YELLOW NICKEL PORPHINOID.

-i- SUBUNIT: HEXAMER OF TWO ALPHA, TWO BETA, AND TWO GAMMA CHAINS.

-i- DEVELOPMENTAL STAGE: THERE ARE TWO MCR COMPLEXES IN THIS BACTERIA. MCR II IS EXPRESSED IN THE EARLY GROWTH PHASE. LATE GROWTH CELLS CONTAINS MOSTLY MCR I.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

EMBL; X07794; CAA30639.1; --  
DR PIR; E28544; E28544.  
DR PDB; 1MRO; 11-NOV-98.  
KW Methanogenesis; Oxidoreductase; Methylation; Multigene family;  
3D-structure.  
FT INIT\_MET 0 0 METHYLATION.  
FT MOD\_RES 256 256 METHYLATION.  
FT MOD\_RES 269 269 METHYLATION.  
SQ SEQUENCE 549 AA; 60379 MW; 979ABB4B6CB8622B CRC64;

Query Match 3.4%; Score 99; DB 1; Length 549;  
Best Local Similarity 19.5%; Pred. No. 1;  
Matches 84; Conservative 56; Mismatches 182; Indels 108; Gaps 18;

QY 141 LRHYQTETH---AMSTYKKEFGPHHSNQWFDVRVWYLSVSPKSPFED-----AYSSGPF 190  
DB 145 VQEHMVETHPALVADSVKVFVTGNDDEIADLPDAFVIDINKQFPDQAEYLKAEVGDIW 204  
QY 191 EFL---TAVSFSEFYVLT---NLLFVPFMSGAAYN---GDMSTVTFGSAQSDSRHML 241  
DB 205 QVVRIPITVSTCDGATTSRWSAMQIGMSMISAYKQAAGEAATGDFAYAKHAEVIMGT 264  
QY 242 GIECIKELLQDDPNVPIVQWIDKWFWRGYRLTLTYAMMDYMPKRV-----MS 292  
DB 265 YLPVRRARGENEPGGVFPF-----GYLADICQSSRVNEDPVRVSLDVVATGML 313  
QY 293 WRSEWMEYAFONGA-----LFFKDLARYG---IREPKQMDACEGKDHIHQ 336  
DB 314 YDQIW-LGYSMSGGVGTQYATAAYTDNILLDDFTYFCKEYVEDKYGLCEAPNNMDTLDV 372  
QY 337 AWS-TTFYGFNAASFHWVPTEDMG-----WLSAKYDPSDFRYRPR 378  
DB 373 ATEVTFYGLEQYEYFALL--EDQFGGSRQAAVVAAGCGTAFATGATGLSGWYLSM 430  
QY 379 FDHWGEQARAGNRYMKTPLMTCQIPMLFTEPGNPTKI-GARESNYLGKHFHFCSDH 437  
DB 431 YLHKEQHSRLG--FYGYDLODQCGASNVFSIRGDEGLPLRLRGPNYPNMYNVGH----- 483  
QY 438 CKDIFDHPQKYQAWLPVHQIHQGNCFPPDADPGAEGFDFLAALDYAVTNGRNLDF 497  
DB 484 -----QGEYAGISQA-----PHAAARGDAFVFNPLV-----KIAFADNLFV 519  
QY 498 DGSDDQKNFA 507  
DB 520 DFTNVRGEFA 529

RESULT 12

641 NPEKNNIQLNYRYANQEYIDQNLGKSANAYQQDIQQVGL--VVGWEIA 687

Db

RESULT 13  
ACE\_HUMAN

ID ACE\_HUMAN STANDARD; PRT; 1306 AA.

AC F12821;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 15-FEB-2000 (Rel. 39, Last annotation update)

DE ANGIOTENSIN-CONVERTING ENZYME, SOMATIC ISOFORM PRECURSOR (EC 3.4.15.1)

DE (ACE) (DIPETIDYL CARBOXYPEPTIDASE I) (KININASE II) (CDI43 ANTIGEN).

GN DCP1 OR DCP OR ACE.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
[1]

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 89071703.

RA Soubrier F., Alhenc-Gelas F., Hubert C., Allegrini J., John M.,  
Tregear G., Corbol P.;  
RAT "Two putative active centers in human angiotensin I-converting enzyme  
revealed by molecular cloning";  
RLE Proc. Natl. Acad. Sci. U.S.A. 85:9386-9390(1988).  
RN [2]

RN PARTIAL SEQUENCE OF 30-46.

RP RP

RX TISSUE=LUNG;

RX MEDLINE; 90110025.

RA Takeuchi K., Shimizu T., Ohishi N., Seyama Y., Takaku F.,  
Yotsumoto H.;  
RAT "Purification of human lung angiotensin-converting enzyme by high-  
performance liquid chromatography: properties and N-terminal amino  
acid sequence.";  
RLE J. Biochem. 106:442-445(1989).  
RN [3]

RN ZINC-BINDING.

RP RP

RX MEDLINE; 91308093.

RA Ehlers M.R., Riordan J.F.;

RAT "Angiotensin-converting enzyme: zinc- and inhibitor-binding  
stoichiometries of the somatic and testis isoymes.";  
RLE Biochemistry 30:7118-7126(1991).

CC -1- FUNCTION: CONVERTS ANGIOTENSIN I TO ANGIOTENSIN II BY RELEASE OF  
THE TERMINAL HIS-LEU, THIS RESULTS IN AN INCREASE OF THE  
VASOCONSTRICTOR ACTIVITY OF ANGIOTENSIN.

CC -1- CATALYTIC ACTIVITY: RELEASE OF A C-TERMINAL DIPEPTIDE,  
OLIGOPEPTIDE-I-XAA-XBB, WHEN XAA IS NOT PRO, AND XBB IS NEITHER  
ASP NOR GLU. CONVERTS ANGIOTENSIN I TO ANGIOTENSIN II.

CC -1- COFACTOR: BINDS TWO ZINC IONS.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- ALTERNATIVE PRODUCTS: THE TESTICULAR ANGIOTENSIN-CONVERTING ENZYME  
IS TRANSCRIBED FROM THE SAME GENE AS THE SOMATIC ISOFORM, PROBABLY  
FROM AN ALTERNATIVE START SITE.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M2 (ZINC METALLOPROTEASE).

CC -1- DATABASE: NAME=PROW; NOTE=CD guide CDI43 entry;  
WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cdi43.htm".  
-----

CC This SWISS-PROT entry is copyright. It is produced through a collabora-  
tion between the Swiss Institute of Bioinformatics and the EMBL Outstation  
at the European Bioinformatics Institute. There are no restrictions on  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commer-  
cial entities requires a license agreement (see http://www.isb-sib.ch/annou-  
nce or send an email to license@isb-sib.ch).

CC -----

DR EMBL; J04144; AAA51684.1; -.

DR PIR; A31759; A31759.

DR PIR; PQ0004; PQ0004.

DR MIM; 106180; -.

DR PFAM; PF01401; Peptidase\_M2; 2.

DR PRINTS; PR00791; PEPTIDPASEA.

DR PROSITE; PS00142; ZINC\_PROTEASE; 2.

DW Hydrolase; Metalloprotease; Carboxypeptidase; Zinc; Dipeptidase;  
KW Glycoprotein; Transmembrane; Duplication; Signal;

```
KW Alternative splicing.
FT SIGNAL 1 29
FT CHAIN 30 1306
FT
FT
FT
FT DOMAIN 30 1259
FT TRANSMEM 1260 1276
FT DOMAIN 1277 1306
FT REPEAT 227 583
FT REPEAT 825 1181
FT METAL 390 390
FT ACT_SITE 391 391
FT METAL 394 394
FT METAL 988 988
FT ACT_SITE 989 989
FT METAL 992 992
FT CARBOHYD 38 38
FT CARBOHYD 54 54
FT CARBOHYD 74 74
FT CARBOHYD 79 79
FT CARBOHYD 111 111
FT CARBOHYD 146 146
FT CARBOHYD 160 160
FT CARBOHYD 272 272
FT CARBOHYD 318 318
FT CARBOHYD 445 445
FT CARBOHYD 509 509
FT CARBOHYD 677 677
FT CARBOHYD 695 695
FT CARBOHYD 714 714
FT CARBOHYD 760 760
FT CARBOHYD 942 942
FT CARBOHYD 1191 1191
FT CONFLICT 35 35
FT CONFLICT 42 42
SQ SEQUENCE 1306 AA; 149714 MW; 1B33BCA7301A26AA CRC64;

Query Match 3.3%; Score 94.5; DB 1; Length 1306;
Best Local Similarity 18.8%; Pred. No. 7.3;
Matches 142; Conservative 80; Mismatches 214; Indels 321; Gaps 46;

Qy 1 MDPTLKKKGLGKRYAMTGLGWETYY---QPMKV--PPDYREGIKHWDKW--- 52
Db 428 VSTPEHLKIGILLR---VTNDESINLLKMALEKIAFLPF---GYLVQW--RWGVF 479
Qy 53 ---VDPRLTMDAYW---KYGEKEKLIYAVIDATONNAFLGVSDARYINALKFLQG 105
Db 480 SGRTPPSRYNFD--WYLLTKYQG---ICPPVTRN---ETHFDGAKFHVN 523
Qy 106 VTP-----LEYLAHRGFARVGRH-----FTGGGARIACQM----- 136
Db 524 VTPYTRYFVSFVLQFPHEALCKEAGYEGPLHQCDIYRSTRKAGAKLRKVLQAGSSRPWQE 583
Qy 137 -----SIDE---LRHYQETHAMSTYKFNFGFHHSNCFWDRVWLVSPKFFEDA 184
Db 584 VLKQMGVGLDALDQLLKYFPQVTLQWLBQNG--QNG--EVLGWPRYQWHPPLP-----DN 635
Qy 185 YSSGPFELT-----AVSFSEFVLNLLFVPMSCA--AYNGDMSTVT----- 226
Db 636 YPEG-IDLVTDAAEKASKVEEYDRISQVWNEYAEANNWNTNITETSKILLQKNMQTA 694
Qy 227 -----FGFSAQS-----DPSRHMTLGIE----- 244
Db 695 NHTLKYGTQARKFDVNLQNTTIKRIKKVODLERAAALPAQBLEEYNKILLDMETYSVA 754
Qy 245 -----CIKELLQDDPNPIVORWIDK--WFWRGYRLLTLVAMMDYMQPKVM--- 291
Db 755 TVCHPNGSLQ--LEPDTNWMATSRKYEDDLWAWEGWRDKAGRAILOFY--PKYVELIN 810
Qy 292 -----SWRESWEMYA--EQNGALFKDL-----ARYGIREPKGWQDAC 327
Db 811 QAARLNGYVDAGDSWRSMTFPTSLQDLERLQELQPLYLNLHAYVVRALHRHYCAQ--- 867
```

```
QY 328 EGDHIS-----HQAWSTFYG-----FNAASAFHTWVPTD--EMGWLSAKY 367
Db 868 ----HINLEGPAPALLGNWQAQTSNIYDLVVPPSPASMDT---TEAMLKOGWTPRRM 920
QY 368 PDSFDRIYR-----PRFDHWG---EQARAGNR-----PY----- 393
Db 921 FKADDDFTSLGLLPVPEF--WNKSMLEKPTDGRREVVCASAWDFYNGKDFRIKQCTTV 978
QY 394 -----MKTPLMLCQTCQIPMLFTEPCGNPTKIGARES----- 424
Db 979 NLEDLVAHHEMCHIQFYMQYKDLPLVAREGANP---GFHEALGDVLALSVTPKHLHSL 1035
QY 425 NYLGNK-----FHCSDHCK-DIFDHE--POKYVQVAMLPVHQ 458
Db 1036 NLLSEGGSDHEDINFLMKWALDKIAFIPFSYLVQWRVDFGSGITKENYQEWWSLRU 1095
QY 459 IHQNCFFP-----DADPGAEGFDPLAAVLDIYYAVT 489
Db 1096 KYQGLC-PPVPRQTQGFDPGAKFHPSSVPYTRYFVS 1131

RESULT 14
AMPN_LACDL STANDARD; PRT; 842 AA.
ID AC P37896;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE AMINOPEPTIDASE N (EC 3.4.11.2) (LYSYL AMINOPEPTIDASE) (LYS-AP)
DE (ALANINE AMINOPEPTIDASE).
GN PEPN.
OS Lactobacillus delbrueckii (subsp. lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 7290;
RX MEDLINE; 94039025.
RA Klein J.R., Klein U., Schad M., Plapp R.;
RT "Cloning, DNA sequence analysis and partial characterization of pepN,
a lysyl aminopeptidase from Lactobacillus delbrueckii ssp. lactis
DSM7290."
RL Eur. J. Biochem. 217:105-114(1993).
CC -!- FUNCTION: AMINOPEPTIDASE WITH BROAD SUBSTRATE SPECIFICITY TO
SEVERAL PEPTIDES. OPTIMUM ACTIVITY IS MEASURED AT PH 6.5-7.0 AND
TEMPERATURES 54-55 DEGREES CELSIUS.
CC -!- COFACTOR: BINDS ONE ZINC ION.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOSOLASMIC. IT MAY BE SECRETED THROUGH
AN UNKNOWN MECHANISM.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1 (ZINC METALLOPROTEASE);
ALSO KNOWN AS THE PEPN SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; Z21701; CAA79805.1; -.
DR PIR; S38364; S38364.
DR PFAM; PF01433; Peptidase M1; 1.
DR PRINTS; PR00756; ALADIPTASE.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Aminopeptidase; Zinc.
FT INIT_MET 0
FT METAL 287 287 BY SIMILARITY.
FT ACT_SITE 288 288 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 291 291 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 310 310 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 374 374 PROTON DONOR (POTENTIAL).
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:24:16 ; Search time 82.1 Seconds  
(without alignments)  
388.947 Million cell updates/sec

Title: US-09-430-029-5  
Perfect score: 2873  
Sequence: 1 MDTPTLKKLGLKDRYAAMT.....FDGSEDKNFAAWRGQATRN 516

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 178050 seqs, 61884766 residues

Total number of hits satisfying chosen parameters: 178050

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_64.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1854.5	64.5	516	1 S44306	phenol 2-monooxyge
2	1846.5	64.3	517	1 D37831	phenol 2-monooxyge
3	1795	62.5	511	1 S47290	phenol 2-monooxyge
4	254.5	8.9	527	2 JQ0702	methane monooxygen
5	211	7.3	525	2 S15207	methane monooxygen
6	115	4.0	852	1 GNLJGA	pol polyprotein -
7	111.5	3.9	852	2 S29358	pol polyprotein - bovi
8	107.5	3.7	852	1 GNLJGB	pol polyprotein -
9	104	3.6	549	1 A24436	alpha-amylase (EC
10	103	3.6	526	2 T01089	hypothetical prote
11	103	3.6	550	2 B69022	methyl coenzyme M
12	99.5	3.5	1032	2 S74487	hypothetical prote
13	99	3.4	550	1 E28544	methyl coenzyme M
14	97	3.4	785	2 H72228	hypothetical prote
15	96.5	3.4	1209	2 T31657	reverse transcript
16	95	3.3	782	1 G64157	probable organic s
17	95	3.3	1336	3 T17479	hypothetical prote
18	95	3.3	1455	1 A48925	mannose receptor p
19	94.5	3.3	1306	1 A31759	peptidyl-dipectida
20	94	3.3	395	2 T45547	arylsulfatase acti
21	94	3.3	452	2 T28094	hypothetical prote
22	93.5	3.3	333	1 S47288	phenol 2-monooxyge
23	93.5	3.3	579	2 JW0071	asparagine synthas
24	92.5	3.2	843	2 S38364	membrane alanyl am
25	91.5	3.2	548	1 ALBSF	alpha-amylase (EC
26	90.5	3.2	362	2 H69785	mannan endo-1,4-be
27	90.5	3.2	1256	1 A43829	muramidase-release
28	90	3.1	549	1 A24549	alpha-amylase (EC
29	89.5	3.1	1258	2 JC5765	inositol polyphosp

30 89 3.1 245 2 T23844  
31 89 3.1 405 2 A35159  
32 89 3.1 671 2 A45730  
33 89 3.1 722 2 B75074  
34 89 3.1 875 2 A36369  
35 88.5 3.1 848 2 T16430  
36 88 3.1 549 1 A54541  
37 88 3.1 583 1 AJPMN2  
38 88 3.1 713 2 B64743  
39 87.5 3.0 561 2 S73087  
40 87.5 3.0 702 2 A34434  
41 87.5 3.0 1009 2 T18533  
42 87.5 3.0 1210 3 T41740  
43 87.5 3.0 2410 2 T43435  
44 87.5 3.0 2410 2 T43731  
45 87.5 3.0 3119 2 T18414

## ALIGNMENTS

## RESULT 1

S44306  
phenol 2-monooxygenase (EC 1.14.13.7) component N - Pseudomonas putida  
N: Alternate names: phenolhydroxylase chain D  
C: Species: Pseudomonas putida  
C: Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C: Accession: E58972; S54765; S44306; S47417  
R: Ng, L.C.; Shingler, V.; Sze, C.C.; Poh, C.L.  
Gene 151, 29-36, 1994  
A: Title: Cloning and sequences of the first eight genes of the chromosomally encoded  
A: Reference number: A58972; MUID: 95129877  
A: Accession: E58972  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-516 <NGL>  
A: Cross-references: EMBL:X79063; NID:9483477; PIDN:CAA55663.1; PID:9483481  
A: Experimental source: strain P35X (NCBI 9869)  
A: Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1994  
R: Herrmann, H.; Mueller, C.; Schmidt, I.; Mahnke, J.; Petruschka, L.; Hahnke, K.  
Mol. Gen. Genet. 247, 240-246, 1995  
A: Title: Localization and organization of phenol degradation genes of Pseudomonas put  
A: Reference number: S54761; MUID: 95272534  
A: Accession: S54765  
A: Status: preliminary; nucleic acid sequence not shown; translation not shown  
A: Molecule type: DNA  
A: Residues: 1-83, 'H', 85-100, 'F', 102-106, 'M', 108-137, 'EL', 140-512, 'Q', 514-516 <HE2>  
A: Cross-references: EMBL:X80765; NID:9527546; PIDN:CAA56743.1; PID:9527550  
A: Experimental source: strain H  
A: Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994  
C: Genetics:  
A: Gene: phdN  
C: Superfamily: phenol 2-monooxygenase component N  
C: Keywords: oxidoreductase

Query Match 64.5%; Score 1854.5; DB 1; Length 516;  
Best Local Similarity 64.8%; Pred. No. 1.1e-142;  
Matches 329; Conservative 65; Mismatches 105; Indels 9; Gaps 1;

QY 5 TLKKLGLKDRYAAMTRGIGLWETTYQPMDKVFPDYREGIKIHDKWDKWDPPRLTMDAYW 64

Db 3 TNKRLNLKDKRYLTROLGWEPYQKKEDVPLEHFEFGIKITDKDKWEDPRLTMDSW 62

QY 65 KYQGEKKLYAVIDAFTONNAFLGVSDARYINALKFLQGVTPLEYLAHGRFAHVGRHF 124

Db 63 KYQAEKKLYAIFDAFQNGHQNHDARYYNALKLELTGVSPLEYQAFQGSRVGRF 122

QY 125 TGEGARIACOMOSIDELHYQETHAMSTPNKFNFGHHSNQWEDRVVWLVKPKSFEDA 184

Db 123 SGAGARVACOMQADIDRVHVQVHAMSHYNKHFQGLDFAHMYRVWFLSVKPKSFMDA 182

QY 185 YSSGPFETAVSFSEFYVLTNLLVFPFMSGAAAYNGDMSTVTFGFSQSDSRHMTLGT 244

Db 183 RTAGPEFLTAVSPSEYVLTNLLFVPMGSAAYNGDMATVTFGSAQSDARHMTLGLE 242  
QY 245 CIKELLQDNDNPIVORWIDKFWRGYRLTLVAMMDYMQPKRVMSWRSWEMAEQN 304  
Db 243 VIKFMLEQHDENVPIIQRWIDKFWRGYRLTLTGMMDYMLPNKVMASWSAWGYPRQA 302  
QY 305 GGALFKDLARYGIREPKGWQDACGKDHISHOAMSTFYGFNAASAFHTWVPTDEMGWLS 364  
Db 303 GGALFKDLARYGIRPPYVQTTTGKHHIHOVWGAFYQYSKATNFHTWIPGDEELNWL 362  
QY 365 AKYPDSFDYRPRFDHWGEQARAGNRFYMTLPMLCOTQIPMLFTPPGNPTKIGARES 424  
Db 363 EKYEDTFDKYRPRFEWREQAQKGERYNDTLPHLCQVCPALFTEPDDPTKLSRSL 422  
QY 425 NYLGNKFHSCDCKDIFDHEPQKYVQAWLPVHQHOGNCFPPDADPGAGFDPPLAAVLID 484  
Db 423 VHEGERYHFCSDGCCDIFKNPEPVKYIOAWLPVHQIYQGN-----EGGDVETVYQK 473  
QY 485 YYAVTMGRDNLDFDGSDDOKNFAAWRGQ 512  
Db 474 YYHKSGVDNLEYLGSPEHQRWLALKGQ 501

## RESULT 2

phenol 2-monooxygenase (EC 1.14.13.7) chain P3 - Pseudomonas sp. (strain CF600)  
C:Species: Pseudomonas sp.

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: D37831

R:Nordlund, I.; Powlowski, J.; Shingler, V.

J. Bacteriol. 172, 6826-6833, 1990

A:Title: Complete nucleotide sequence and polypeptide analysis of multicomponent phenol

A:Reference number: A37831; MUID:91072230

A:Accession: D37831

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-517 <NOR>

A:Cross-references: GB:M60276; GB:M37764; NID:gl51449; PID:AAA25942.1; PID:gl51453

C:Superfamily: phenol 2-monooxygenase component N

C:Keywords: oxidoreductase

Query Match 64.3%; Score 1846.5; DB 1; Length 517;  
Best Local Similarity 64.8%; pred. No. 4.8e-142;  
Matches 328; Conservative 63; Mismatches 106; Indels 9; Gaps 1;

QY 7 KKKLGLKDRYAAMTGLGWETTYQPMCKVPPDYRYEGIKIHWDKWDVPPRLTMDAYWKY 66  
Db 6 KKRNLKDKYRLTRDAWETTYOKKEDVPLEHFEGIKITDWDKWDVPPRLTMDTYWKY 65  
QY 67 QGEKKLYAVIDAFTONNAFLGVSADARYNALKLFQGVTPLEYLAHRGFARHFTG 126  
Db 66 QAEKKLYAIFDAFQNNQGNISADARYNALKLFITAVSPLEYQAFQGSFVRGQFSG 125  
QY 127 EGARIACQMOSIDELRHQYQETHAMSTYKPNFNGFHHSNQWFDVWYLSVPKSFEDAYS 186  
Db 126 AGARVACQMAIDELRHVQTVHAMSHYKHFQGLHDFAHMDYDRVWYLSVPKSYMDART 185  
QY 187 SGPEFTTAVSFEYVLTNLLFVPMGSAAYNGDMSTVTFGSAQSDSHRMTLGIECI 246  
Db 186 AGPEFTTAVSFEYVLTNLLFVPMGSAAYNGDMATVTFGSAQSDARHMTLGLEVI 245  
QY 247 KFLLEQDNDNPIVORWIDKFWRGYRLTLVAMMDYMQPKRVMSWRSWEMAEQNGG 306  
Db 246 KFMLEQHDENVPIIQRWIDKFWRGYRLTLTGMMDYMLPNKVMASWSAWGYPRQA 305  
QY 307 ALFKDLARYGIREPKGWQDACGKDHISHOAMSTFYGFNAASAFHTWVPTDEMGWLSAK 366  
Db 306 ALFKDLARYGIRPPYVQTTTGKHHIHOVWGAFYQYSKATNFHTWIPGDEELNWLSEK 365  
QY 367 YPDSFDYRPRFDHWGEQARAGNRFYMTLPMLCOTQIPMLFTPEGNPTKIGARES 426  
Db 366 YPDSFDYRPRFDHWGEQARAGNRFYMTLPMLCOTQIPMLFTPEGNPTKIGARES 426

Db 366 YPDTFDKYRPRFEWREQAQKGERFYNDTLPHLCQVQCLPVITFTEPDDPTKLSRSLVH 425  
QY 427 LGNKFHSCDCKDIFDHEPQKYVQAWLPVHQHOGNCFPPDADPGAGFDPPLAAVLIDY 486  
Db 426 EGERYHFCSDGCCDIFKNPEPVKYIOAWLPVHQIYQGN-----EGGDVETVYQY 476  
QY 487 AVTMGRDNLDFDGSDDOKNFAAWRGQ 512  
Db 477 HKSGVDNLEYLGSPEHQRWLALKGQ 502

## RESULT 3

S47290

phenol 2-monooxygenase (EC 1.14.13.7) chain mopN - Acinetobacter calcoaceticus

N:Alternate names: phenol hydroxylase

C:Species: Acinetobacter calcoaceticus

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: S70083; S47290

R:Ehrt, S.; Schirmer, F.; Hillen, W.

Mol. Microbiol. 18, 13-20, 1995

A:Title: Genetic organization, nucleotide sequence and regulation of expression of ge

A:Reference number: S70080; MUID:96154937

A:Accession: S70083

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-511 <EH2>

A:Cross-references: EMBL:Z36909; NID:9535279; PID:CAA85383.1; PID:9535283

A:Experimental source: strain NCIB8250

A>Note: The nucleotide sequence was submitted to the EMBL Data Library, September 199

C:Genetics:

C:Superfamily: phenol 2-monooxygenase component N

C:Keywords: aromatic hydrocarbon catabolism; oxidoreductase

Query Match 62.5%; Score 1795; DB 1; Length 511;  
Best Local Similarity 64.1%; pred. No. 7.1e-138;  
Matches 325; Conservative 60; Mismatches 112; Indels 10; Gaps 2;

QY 8 KKLGLKDRYAAMTGLGWETTYQPMCKVPPDYRYEGIKIHWDKWDVPPRLTMDAYWKY 67  
Db 13 KKLNAKERYRLTRDLWDWDFSYADKDAFPYEEFEKIKITDWSKWEDPFRITMDNYWKY 72  
QY 68 GEKKLYAVIDAFTONNAFLGVSADARYNALKLFQGVTPLEYLAHRGFARHFTG 127  
Db 73 AEKKLYAIFDAFQNNQGNISADARYNALKLFITAVPLEYQAFQGSFVRGQFSGI 132  
QY 128 GARIACQMOSIDELRHQYQETHAMSTYKPNFNGFHHSNQWFDVWYLSVPKSFEDAYS 187  
Db 133 GARIACQMOSIDELRHQYQETHAMSHYKHFQGLHDFAHMDYDRVWYLSVPKSFEDAKSA 192  
QY 188 GPFPELTAVSFEYVLTNLLFVPMGSAAYNGDMSTVTFGSAQSDSHRMTLGIECIK 247  
Db 193 GPFPELTAVSFEYVLTNLLFVPMGSAAYNGDMATVTFGSAQSDARHMTLGLEIVK 252  
QY 248 FLLEQDNDNPIVORWIDKFWRGYRLTLVAMMDYMQPKRVMSWRSWEMAEQNGGA 307  
Db 253 FLLEQHDENVPIVQEWIDKFWRGYRLTLTGMMDYMLPNKVMASWSAWGYPRQA 312  
QY 308 LFKDLARYGIREPKGWQDACGKDHISHOAMSTFYGFNAASAFHTWVPTDEMGWLSAKY 367  
Db 313 LFKDLARYGIRPPYVQTTTGKHHIHOVWGAFYQYSKATNFHTWIPGDEELNWLSEK 372  
QY 368 YPDSFDYRPRFDHWGEQARAGNRFYMTLPMLCOTQIPMLFTPEGNPTKIGARES 426  
Db 373 YPDSFDYRPRFDHWGEQARAGNRFYMTLPMLCOTQIPMLFTPEGNPTKIGARES 432  
QY 427 LGNKFHSCDCKDIFDHEPQKYVQAWLPVHQHOGNCFPPDADPGAGFDPPLAAVLIDY 486  
Db 433 KDERYHFCSDGCCDIFKNPEPVKYIOAWLPVHQIYQGN-----EGGDVETVYQY 483  
QY 487 AVTMGRDNLDFDGSDDOKNFAAWRGQ 512



Db 484 NFNVGADNLDIEGSPDOORWKKWGNA 510

RESULT 4

QJ00702

methane monooxygenase (EC 1.14.13.25) component A alpha chain - Methylococcus capsulatus

C;Species: Methylococcus capsulatus

C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 18-Jun-1999

C;Accession: JQ0702

R;Stainthorpe, A.C.; Lees, V.; Salmond, G.P.C.; Dalton, H.; Murrell, J.C.

Gene 91, 27-34, 1990

A;Title: The methane monooxygenase gene cluster of Methylococcus capsulatus (Bath).

A;Reference number: JQ0700; MUID:90382694

A;Accession: JQ0702

A;Molecule type: DNA

A;Residues: 1-527 <STA>

C;Comment: This multicomponent enzyme catalyzes the conversion of methane to methanol us

C;Genetics:

A;Gene: mmox

C;Keywords: oxidoreductase

Query Match 8.9%; Score 254.5; DB 2; Length 527;

Best Local Similarity 24.2%; Pred. No. 6.1e-13;

Matches 109; Conservative 65; Mismatches 179; Indels 97; Gaps 22;

Qy 54 DPFLRTMDAYWKYQGEKEKKLYAVID-AFTONNAFLGVSDARYINALKLFLOGVTPLEVL 112

Db 58 EBFKLIARKEAYARMAVKDERQFSGISQVALTRLNAGRV-HFKWNETMKVV----- 106

Qy 113 AHRGFAHVGRH-----FTG---EGARIAQ-----MQSIDELRHQTEPHAMSTYNKFF-- 158

Db 107 --SNFLEGEYNATAATGMLWDSQAABQKNGYLAQVLDELIRH-----THQCAIYVNYEAK 160

Qy 159 NG---FHHSNWFDR---VWVLSVPKSFEDAYSSGCFELTAVSFSE-----EYVLT 205

Db 161 NGQDPAGHNDARRRTTGPLW--KGMKRVFSDGFLSG-----DAVECSLNLQLVGEACFT 213

Qy 206 NLLFVPFMSGAAINGDMSTVTFGSAQSDSRHMTLGTGECIKFLLEQDPDNPVIVORWID 265

Db 214 NPLIVATWEAANGDEIIPVFISIEIDELRHMANGYQTV-VSIANDPASAKYINTDNLN 272

Qy 266 KWFWRGVRLLT-LVAMMDYMPKRVMSWRSEWEMYA-EQNGGALFKDLARYGIREPKGW 323

Db 273 NAFVTQKIFYTPVLGLMFEYGSKEKVEPWKTNRWVYEDWGGIWIIGRLGYGVESPRSL 332

Qy 324 QDACEGDHISHQAWSTFYGFNAASAFHTWVPTEDENGWLAKYPDSFDRIYRFRFDHWG 383

Db 333 KDAQDAYWAHHDYLLAYALWPTGFFRLALPDQEEEMEFANYPGWYD-HYGKIYESEW- 390

Qy 384 EQARAGNRFRYMKTLPLMLCQTQCIQ-----MLFTEPCNPTKI-----CA- 421

Db 391 -RARG-----CEDSSGFILPMWFIENNHPYIDRVSVQVPCPSLAKGAS 434

Qy 422 --RESNYLGNKHFCHSDCKDIFDHEPOKY 449

Db 435 TLRVHEYNGQMTSFSDQWGERMMLAEPEPY 464

RESULT 5

SI5207

methane monooxygenase (EC 1.14.13.25) component A alpha chain - Methylosinus trichospori

C;Species: Methylosinus trichosporium

C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 18-Jun-1999

C;Accession: SI5207; A39049

R;Cardy, D.I.N.; Laidler, V.; Salmond, G.P.C.; Murrell, J.C.

Mol. Microbiol. 5, 335-342, 1991

A;Title: Molecular analysis of the methane monooxygenase (MMO) gene cluster of Methylosi

A;Reference number: SI5207; MUID:91251762

A;Accession: SI5207

A;Molecule type: DNA

A;Residues: 1-525 <CAR>

A;Cross-references: EMBL:X55394; NID:g44613; PIDN:CAA39068.1; PID:g44614

R;FOX, B.G.; Liu, Y.; Dege, J.E.; Lipscomb, J.D.

J. Biol. Chem. 266, 540-550, 1991

A;Title: Complex formation between the protein components of methane monooxygenase fr

A;Reference number: A39049; MUID:91093180

A;Accession: A39049

A;Status: preliminary

A;Molecule type: protein

A;Residues: 2-16 <FOX>

C;Genetics:

A;Gene: mmox

C;Keywords: oxidoreductase

Query Match 7.3%; Score 211; DB 2; Length 525;

Best Local Similarity 21.0%; Pred. No. 2.1e-09;

Matches 96; Conservative 71; Mismatches 196; Indels 94; Gaps 19;

Qy 45 KIHWDKWDVDFRLTMDAYWKYQGEKEKKLY-AVIDAFTQNNA----- 86

Db 49 KYHMANETKEQFKVIAKEYARMEAAKDERQFGTLLDGLTRLGAGNKVHPRWGETMKVTSN 108

Qy 87 FLGVSDARYINALKLFLOGVTPLE-----YLAHRGFAHVGRHETGEGARIACOMQSIDEIR 142

Db 109 FLEVGEYNALIAASAMLWDSATAAEQKNGYLA-----QVLDEIR 146

Qy 143 HYQTEHAMSTYNKPFENGFFHS-----NQWFDVWVLSVPKSFEDAY-SSGPTEF 192

Db 147 H-----THQCAFINHYYSKHYHDHPAGHNDARRTRATGPLW--KGMKRVFADGFISSDAVBC 200

Qy 193 LTAVSFSEFYVLNLLFVPFMSGAAINGDMSTVTFGSAQSDSRHMTLGTGECIKFLLEQ 252

Db 201 SVNQLQVGDTCFTNPLIVAVTEWALGNDEITPTVFLSVETDELHRMANGYQTV-VSIAN 259

Qy 253 DPNVPIVQVRWIDKFWRGYRLIT-LVAMMDYMPKRVMSWRSEWEMYAEGN-GGALFK 310

Db 260 DPASAKFLNTDLNNAFTQKQYFTFVLGYLFEYGSKEKVEPWKTNRWVYEDWGGIWI 319

Qy 311 DLARYGIREPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTW-----VPTEDMGW 362

Db 320 RLKYGIV-ESRVLRAKRDADYWAHHDL-----ALAYAMWPLAFARLALPDEEDQAW 370

Qy 363 LSAKYPSDFRYYRPRFDHWG---EQARAGNRFRYMKTIP-----MLQQTQCIQPLMFTPE 413

Db 371 FEANYPGWAD-HYGKIFNEWKGLGYEDPKSGFIPYQWLLANGHDYVIDRVSVQVFI---P 426

Qy 414 GNPIKIGA-RSNNYLGKHFHFCSDHCKDIFDHEPOKY 449

Db 427 SLAKGTSLRVHFEFGNKKHSLTDDWGERQWLLEPERY 463

RESULT 6

GNLJGA

pol polyprotein - bovine leukemia virus (strain Australia)

N;Contains: endonuclease (EC 3.1.-.-); RNA-directed DNA polymerase (EC 2.7.7.49)

C;Species: bovine leukemia virus, BLV

A;Note: host Bos sp. (cattle)

C;Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 16-Jul-1999

C;Accession: JQ0555

R;Coulston, J.; Naif, H.; Brandon, R.; Kumar, S.; Khan, S.; Daniel, R.C.W.; Lavin, M.

J. Gen. Virol. 71, 1737-1746, 1990

A;Title: Molecular cloning and sequencing of an Australian isolate of proviral bovine

A;Reference number: JQ0554; MUID:90362060

A;Accession: JQ0555

A;Molecule type: DNA

A;Residues: 1-852 <COO>

A;Cross-references: DBJ:D00647; NID:g221051; PIDN:BAA00544.1; PID:g1000999; PID:g221

A;Note: This reading frame extends between two stop codons and does not begin with a

A;Note: the authors translated the codon CCC for residue 514 as Gln

C;Comment: The precise boundary between RNA-directed DNA polymerase and endonuclease

C;Genetics:

A;Gene: pol

C;Superfamily: pol polyprotein

C;Keywords: endonuclease; hydrolase; nucleotidyltransferase; polyprotein; reverse tra



Best Local Similarity 19.0%; Pred. No. 0.94;  
Matches 115; Conservative 70; Mismatches 214; Indels 207; Gaps 30;

```
QY 3 TPTLKKGLKDRYAAMTGLGWET-----TYQPM-----DKVFPYDRYGEIKTH 47
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 169 STEBORSOCYALAAARLDLGFQVASEKTSQPSVPFLGQWVHEQIVTYQSLPTLQIS 228
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 48 DWDKWWDPFRLTW-----DAYWKYQGEKEK-----LYAVIDAFTQNNAFGLSDARYI 96
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 229 -----SPISLHQLQAVLDQLQWVRGTPTRRPLQLLYSSLKRHHDPRAIIQLSEQ-- 280
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 97 NALKLELQGVTL- EYLARG-----FAHVGRRHTGCGARTAC-OMOSIDELRIY 144
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 281 -----LOGIAELRQALSHNARSRYNEQEPFLAYV--HLTRAGSTLVLFQKGAQFPFLAYF 332
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 145 QT-----ETHAMSTYNKFFNGFRHS-----NOMED----- 169
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 333 QPFLTDNQSPWCLLLLLGCQYLQALSSYAKPIKLYHNLPKISLDNWIOSSEDPVQ 392
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 170 -----RVWYLSVPKSFTEDAYSSGPFE-----FLTAVSFSEFYVLTNLLFVPMFMSGAAY 218
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 393 ELLQLW-----POISSOGIQPPGPWKLTITRAEVLTP-QFSPDPIPAALCL--FSDGATG 445
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 219 NG-----DMSTVTFEGSAQSDSRHMTLGTICIKELLEDQDPNDVPIVORWID-KW 267
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 446 RGAYCLWKDHLDFQAVPAPESAQKGLAGLAG-----LAAAPPEPVNI---WYDSKY 496
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 268 FWRGYRLTLVAMMDYMPKRVMSNR-----WLOQDPVPSVALLYKSLLRHPAIVGHRVSHSASHPIASLNNY 552
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 497 LKSLRTLVLGA-----ESWEMVAEQNGALFK-----DLARYGIREPKGWQDACE-----GKDHI 333
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 295 -----ESWEMVAEQNGALFK-----DLARYGIREPKGWQDACE-----GKDHI 333
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 553 VDOLLPLETPEQWHKLTCHNSRALSNWPNRISAWDRSPATLCEFCQKLNPTGGCKMRT 612
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 334 SIQAWS-----TFYGFNAAS-AFTWVTEDEMGWLSAKYDPSDFRYRPRFD--- 380
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 613 IQRGNAPNHIAQADITHYKYKFTYALHVFVDITYSGATHASAKRGLTQTOTIEGLEIAIV 672
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 381 HWGEQAR-----AGNRFYMTLPMQLCQTCOIPMLFTEPGNPTKIGARES-----NYLGKNF 431
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 673 HLCGRPKLNTDOGANYSKTFRFCQGFVSHHVPYNTSSGLDERINGLLKLLLSKY 732
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 432 HPCSDH 437
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 733 HLDEPH 738
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
```

```
RESULT 9
A24436
alpha-amylase (EC 3.2.1.1) precursor - Bacillus stearothermophilus plasmid pat5
N:Alternate names: 1,4-alpha-D-glucan glucanohydrolase
C:Species: Bacillus stearothermophilus
C>Date: 05-Jun-1987 #sequence_revision 18-Aug-1995 #text_change 18-Jun-1999
C:Accession: A24436; 139777
R:Nakajima, R.; Imanaka, T.; Alba, S.
J. Bacteriol. 163, 401-406, 1985
A:Reference number: A24436; MUID:85234394
A:Accession: A24436
A:Molecule type: DNA
A:Residues: 1-349 <NAK>
A:Cross-references: GB:M1450
A:Experimental source: plasmid pat5
A>Note: amino end of the mature protein also determined
R:Jorgensen, P.L.; Hansen, C.K.; Poulsen, G.B.; Diderichsen, B.
Gene 96, 37-41, 1990
A:Title: In vivo genetic engineering: homologous recombination as a tool for plasmid con
A:Reference number: 139772; MUID:91092499
A:Accession: 139777
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-45 <RES>
A:Cross-references: GB:M62638; NID:g142514; PIDN:AAA22242.1; PID:g142515
```

C:Comment: Alpha-amylase genes have been found on plasmids and in multiple copies on  
C:Genetics:  
A:Gene: amys  
A:Genome: plasmid  
A:Start codon: GTG  
C:Function:  
A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds  
A:Pathway: glycogen/starch degradation  
C:Superfamily: alpha-amylase, amyloliquefaciens type; alpha-amylase core homology  
C:Keywords: extracellular protein; glycosidase; glycosidase; heat-stable protein; hydrolase; polys  
F:1-34/Domain: signal sequence #status predicted <SIG>  
F:35-549/Product: alpha-amylase #status experimental <MAT>  
F:235-368/Domain: alpha-amylase core homology <AMY>  
F:139,237,272/Binding site: calcium (Asp, Asp, His)  
F:268,298,365/Active site: Asp, Glu, Asp #status predicted

Query Match 3.6%; Score 104; DB 1; Length 549;  
Best Local Similarity 18.3%; Pred. No. 1;  
Matches 82; Conservative 43; Mismatches 125; Indels 198; Gaps 24;

```
QY 21 RGLG--WEITYQPMKVFYDRYEGT-----KIHWDKW-----VDPELTMDAY 63
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 213 RGIGKAWDEVPDENGNYDLYMADLMDHPDEVVTELKSGWKYVNTNIDGFRLDVAKH 272
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 64 WKYQGEKEKKLYAVIDAFTQNNAF--GVSDARYINALKLELQGVTPLEYLAHRGFAHVG 121
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 273 IKF-----SFFPDWLSQVRSQTKPLFTVG-----EWSY----- 302
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 122 RHPTGEGARACOMQSIDELRHRYQTEHAMSTYNKFFNGPHHSNQWDFRVWYLSVPKSF 181
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 303 -----DINKLHNYIMKTN--GTMSLFDAPLN-----KFY 330
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 182 EDAYSSCPPEFLTAVSFSEFYVLTNLLFVPMFMSGAAYNGDMSTVTFGFSQSDSRHML 241
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 331 TASKSGGTDMRT-----LMNTLM-----KDQPTL----- 356
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 242 GICIKFLELQDPNDVPIVORWIDKWF-----NRGYRL----- 275
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 357 ---AVTFVDNHDEPQALQSWDPWPKPLAYAFILTRQEGYPCVFGYGYGIPQYNIPS 413
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 276 -----TLVA-----MMDYMQPKRVMSW-RE-----SWEM 299
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 414 LKSKIDPLLIARDYAYGTOHDYLDHSDIIGTWREGTEKPGSGLAALITDGGSKWMY 473
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 300 YAEONGALFKDLARYGIREPKGWQDACEGKDHISHQAWSTFFYGFNAASAFHTWVP---T 356
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 474 VGQKHAGKVFDLT--GNR-----SDTVT-----INSDGWGEF-KVNGGSV-SYMWVPRKTT 520
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 357 EDEMGWLSAKYP--DSFDRYRPRFDHW 382
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 521 VSIANSITRPTWDEFWTEFLVAV 548
   :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
```

RESULT 10  
T01089  
hypotheical protein T10P11.12 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 14-May-1999  
C:Accession: T01089  
R:Kaplan, N.; Johnson, D.; Schutz, K.; Gnoj, L.; Hoffmann, J.; Till, S.; de la Bastide  
hi, M.; Martienssen, R.; Chen, E.Y.; Wilson, R.; McCombie, W.R.  
submitted to the EMBL Data Library, November 1998  
A:Description: Sequence of A. thaliana BAC T10P11 from chromosome IV.  
A:Reference number: Z14248  
A:Accession: T01089  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-526 <KAP>  
A:Cross-references: EMBL:AC002330; NID:g2262135; PID:g3892049  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 4

A;introns: 61/1; 145/3; 184/3; 205/1; 237/3; 262/3; 309/1; 322/3; 392/3; 415/2; 435/3  
A;Note: T10P11.12

```
Query Match          3.6%; Score 103; DB 2; Length 526;
Best Local Similarity 22.0%; Pred. No. 1.2;
Matches 71; Conservative 43; Mismatches 130; Indels 78; Gaps 17;

QY 45 KIHWDKWDVDFRLTMDAYKYGQF---KEKKLYAVIDAFTONNAFLGVSDARYI----- 96
Db 184 RIHQWKWEDSI-----ADEKFDDETALRRKRVTHVHNHAFTEKH-FIGICKDSVILGWT 237
QY 97 -NALKFLQGYTPLEYLAHR-GFARVGRHFTGEGARIACQMSIDELRHYOTETHAMSTY 154
Db 238 QSEFLKQFVDSVTKSDYVTLRLGF--IWHCKG-NPKLNFHYMMRALEDDEKQVGVISWY 294
QY 155 NKFF-----NGFHHSNOWERVWVLSVPKSFEDFAYSSGPFETAVSFSFEYVLT 205
Db 295 LWIEFVFIPLLNVGWH-----TYFWIAFT--PFALLAVGTKLEHVIA 336
QY 206 NLLFVPMMSGAAVNGDM---STVTFGSAQSDSRHMTLGIECTKFLLEDPPONVPIVQ 261
Db 337 QLAHEVAEKHVAIEDGLVWPKSDHFEFSKPQ-----IVYLIHLFLFQNAFEIAFF- 388
QY 262 RWIDKFWRGYRLTLVAMMMYMQPK-----RVMSWRESWEMYA--EONGG-----AL 308
Db 389 ----FWIWTYVGEDSCIMGQVRYIVPRVIGVFIQVLCYSTPLPLAIYVSMQSGSFKKAI 444
QY 309 FKDLARYGIREPKWQDACEGK 330
Db 445 FEENVQVGL---VGWAQVKQK 463

RESULT 11
B69022
methyl coenzyme M reductase (EC 1.8.-.-) I alpha chain - Methanobacterium thermoautotroph
C;Species: Methanobacterium thermoautotrophicum
A;Variety: Strain Delta H
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jun-1999
C;Accession: B69022; S78579
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.
; K.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514
A;Accession: B69022
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-550 <MTH>
A;Cross-references: GB:AE000885; GB:AF000666; NID:g2622256; PIDN:AAB85653.1; PID:g262227
A;Experimental source: strain Delta H
R;Rospert, S.; Linder, D.; Ellermann, J.; Thauer, R.K.
Eur. J. Biochem. 194, 871-877, 1990
A;Title: Two genetically distinct methyl-coenzyme M reductases in Methanobacterium therm
A;Reference number: S13864; MUID:91099370
A;Accession: S78579
A;Molecule type: protein
A;Residues: 2,'X',4-15 <ROS>
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH164
C;Complex: heterohexamer of two alpha, two beta (see PIR:F69022) and two gamma (see PIR:
C;Function:
A;Description: catalyzes the reaction of methyl coenzyme M (2-(methylthio)enthanesulfoni
de and methane
A;Pathway: methanogenesis
C;Superfamily: methyl coenzyme M reductase alpha chain
C;Keywords: heterohexamer; metalloprotein; methanogenesis; nickel; oxidoreductase
F;2-550/Product: methyl coenzyme M reductase I alpha chain #status experimental <MAT>
F;147/Binding site: coenzyme F430 nickel (Gln) (axial ligand) #status predicted
F;333/Active site: Tyr #status predicted
F;444/Binding site: coenzyme M (Tyr) #status predicted
```

F;481/Binding site: coenzyme B (Asn) #status predicted

```
Query Match          3.6%; Score 103; DB 2; Length 550;
Best Local Similarity 19.5%; Pred. No. 1.3;
Matches 85; Conservative 54; Mismatches 178; Indels 118; Gaps 17;

QY 141 LRHYOTETH---AMSTYKFNFGPHHSNOWERVWVLSVPKSFEDF-----AYSSGPF 190
Db 146 VOEHMVEITHPALVADSYVKVFTGNDIEADEIDPAEVIDINKOFFPEDQAFETKAEVGDGIW 205
QY 191 EFL---TAVSFSEFYVLT---NLLFVPMMSGAAVY---GDMSTVTFGSAQSDSRHMTL 241
Db 206 QVVRIPITVSRCDGATTSRWSAMQIGMSISAYKQAAGEAATGDFAYAAKHAIEVHMG 265
QY 242 GIECKFLLEDPPDNVPIVQRWIDKFWRGYRLTLVAMMMYMQPKRY-----MS 292
Db 266 YLPVVRARGENEPGGVPF-----GYLADICQSSRVNIEDFVRVSLDVVATGAML 314
QY 293 WRESWEMVAEONGGALFKDLA-----RYGIREPKWQDACEGK 330
Db 315 YDQIW-LGSYMSGGVGFQYQYATAATDNLDDFTYFGKEYVEDKIGCEAPNTMTV--- 370
QY 331 DHISHQAWSTFYGFNAASAFHTWPTDEMG-----WLSAKYPDSEDR 373
Db 371 --LDVASEVTFYGLQEYEEYPALL--EDQFGSGQRAAVVAAAAGCSTAFATANAQTGLSG 426
QY 374 YRPRFDHWGQARAGNRFYMKTLPMLCQTQIPMLFTEPGNPTKI-GARSNYLGKFKH 432
Db 427 WYLSMYLHKHQHSRLG--FYGYDLQDQCGASNVFSIRGDEGLPLEIRGPNPNYAMNYGH 484
QY 433 FCSDHCKDIEDHEPKYVQAWLPVHQIHQGNCFPPDADPGAGGPDPLAALVDIYAVTNGR 492
Db 485 -----QGEYAGISQA-----PHARGDAFVFNPLV-----KIAFAD 515

RESULT 12
S74487
hypothetical protein sl11060 - Synechocystis sp. (strain PCC 6803;
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S74487
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S74487
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1032 <KAN>
A;Cross-references: EMBL:D90899; GB:AB001339; NID:gl651650; PIDN:BAAL6639.1; PID:d101
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

Query Match          3.5%; Score 99.5; DB 2; Length 1032;
Best Local Similarity 20.5%; Pred. No. 5.3;
Matches 58; Conservative 50; Mismatches 100; Indels 75; Gaps 15;

QY 54 DPFRLTMDAYKYGQEKKLYAVIDAFTONNAF-----LGVSDA-----RYI-NALKL 101
Db 712 DPLYLWTA---KVGGEHSTLYWLIDAYTTSNSYPSDPGEGDANQPGRNENYIRNSVKI 767
QY 102 FLQGVY-----PLEYLAHRGFAHGRHPTGGARIACQMSIDELRHVTE-- 147
```

```
Db 768 LVDAYNGDVREFFTDKODPLINAWQKIFPELFPSGMPSTLKSHIRYPVDMFSTQSERL 827
QY 148 -THAMSTYNKFFNGFHSNQWDFRWWVYLSVPKSPFFEDAYSS-GPFEF---LTAVSFSEFY 202
Db 828 LTYHMEDIDVFN--REDQW-----RPOEYIADQOPIAPYILIMKLADAKEE 877
QY 203 VLNLFLVPPMSGAAYNGDMSTVTFGSAQSDSRH-----MTLGICIKELL 250
Db 878 FVLSQVTP-----NARNNLIALLE-ARCEQNYGKLLLYTLPKERLVYGPQIEALV 929
QY 251 EQDPDNVPIQVORWIDKWFWRGYR-----LLTLVAMMDYMQP 287
Db 930 NQD-----PVISERLSLWNRGRSRAIQGNLLVPIEESLLYVEP 968

RESULT 13
E28544
Methyl coenzyme M reductase (EC 1.8.-.-) I alpha chain - Methanobacterium thermoautotroph
C:Species: Methanobacterium thermoautotrophicum
A:Variety: strain Marburg
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 15-Oct-1999
C:Accession: E28544; S78576
R:Bokranz, M.; Baeumner, G.; Allmansberger, R.; Aukel-Fuchs, D.; Klein, A.
J. Bacteriol. 170, 568-577, 1988
A:Title: Cloning and characterization of the methyl coenzyme M reductase genes from Meth
A:Reference number: A91891; MUID:88115150
A:Accession: E28544
A:Molecule type: DNA
A:Residues: 1-550 <BOK>
A:Cross-references: GB:X07794; GB:M18969; NTD:944607; PIDN:CAA30639.1; PID:g44612
A:Experimental source: strain Marburg
R:Rospert, S.; Linder, D.; Ellermann, J.; Thauer, R.K.
Eur. J. Biochem. 194, 871-877, 1990
A:Title: Two genetically distinct methyl-coenzyme M reductases in Methanobacterium therm
A:Reference number: S13864; MUID:91099370
A:Accession: S78576
A:Molecule type: protein
A:Residues: 2-19 <ROS>
A:Experimental source: strain Marburg
R:Ermler, U.; Grabarse, W.; Shima, S.; Goubeaud, M.; Thauer, R.K.
Science 278, 1457-1462, 1997
A:Title: Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological
A:Reference number: A58866; MUID:98035783
A:Contents: annotation; X-ray crystallography, 1.45 angstroms
A:Note: the modified residue designated as N1-methylhistidine is correctly named 3'-meth
C:Complex: heterohexamer of two alpha, two beta (see PIR:A28544) and two gamma (see PIR:
C:Function:
A:Description: catalyzes the reaction of methyl coenzyme M (2-(methylthio)ethanesulfoni
de and methane
A:Pathway: methanogenesis
C:Superfamily: methyl coenzyme M reductase alpha chain
F:Keywords: heterohexamer; metalloprotein; methanogenesis; methylated amino acid; nickel
F:2-550/Product: methyl coenzyme M reductase I alpha chain #status experimental <NAT>
F:447/Binding site: coenzyme F430 nickel (Gln) (axial ligand) #status experimental
F:257/Modified site: 3'-methylhistidine (His) #status experimental
F:271/Modified site: 5-methylarginine (Arg) #status experimental
F:333/Active site: Tyr #status predicted
F:400/Modified site: 2-methylglutamine (Gln) #status experimental
F:444/Binding site: coenzyme M (Iyr) #status experimental
F:445/Modified site: 1-thioglycine (Gly) #status experimental
F:452/Binding site: methyl (Cys) (covalent) #status experimental
F:481/Binding site: coenzyme B (Asn) #status experimental

Query Match 3.4%; Score 99; DB 1; Length 550;
Best Local Similarity 19.5%; Pred. No. 2.6;
Matches 84; Conservative 56; Mismatches 182; Indels 108; Gaps 18;

QY 141 LRHYQTETH---AMSTYNKFPNGFHSNQWDFRWWYLSVPKSPFFED-----AYSSGPF 190
Db 146 VQEHMVETHPALVADSVKVTGNDETADEIDPAFVIDINKQFPEDQQAETLKAEVGGIIV 205
QY 191 EFL---TAVSFSEFYVLT---NLLFVPFMSGAAYN---GDMSTVTFGSAQSDSRHMTL 241
```

```
Db 206 QVVRIPITVSRCTDGAATTSRWSAMQIGMSMISAYKQAAGEAAIGDFAYAAKHAIEVHMGT 265
QY 242 GIECIKELLEODPNVPIQVORWIDKWFWRGYRLLTLVAMMDYMQPKRV-----MS 292
Db 266 YLPVRRARGENEPGGVPF-----GYLADICQSSRVNYEDPVKVSVDVVAIGAML 314
QY 293 WRESWENYAONGGA-----LFKDLARYG---IREPKGWQDACEGKDHISHQ 336
Db 315 YDQIW-LGSYMSGVGFTQYATAATYDNLDDFTYFGKEYVEDKYGLCEAPNNMDTVLDV 373
QY 337 AWS-TFYGFNAASAPHWVPTEDMG-----WLSAKYPSDFORYRPR 378
Db 374 ATEVTFYGLEQYEYEPALL--EDQFGGSQRAAVVAAGACSTAFATGNAQTGLSGWLSM 431
QY 379 FDHWGEQARGNRFYMKTLPMLCOTCQIPMLFTEPGNPTKI-GAESNYLNKNCHEFCSDH 437
Db 432 YLHKEQSRLLG--FYGYDLQOCGASNVFSRGGEGLEPLRGNPNYAMNVGH----- 484
QY 438 CKDIFDHPQKYVQAWLFPVHQIHQGNCFPPDADPGAEGFDPLAAVLDYVAVTMGRDNLDF 497
Db 485 -----QGEYAGISQA-----PHAAGDAFVENPLV-----KIAFADDDLNVF 520
QY 498 DGSDEQKNFA 507
Db 521 DFTNVRGEFA 530

RESULT 14
H72228
hypothetical protein TM1624 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: H72228
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: H72228
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-785 <ARN>
A:Cross-references: GB:AF001806; GB:AE000512; NID:g4982196; PID:g4982197; TIGR:TM1624
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1624

Query Match 3.4%; Score 97; DB 2; Length 785;
Best Local Similarity 17.8%; Pred. No. 6;
Matches 86; Conservative 68; Mismatches 147; Indels 190; Gaps 28;

QY 54 DPFRLTMDAYWKY-----QGEKEKKIYAVIDAFTQ---NNAFLGVSDARYI--- 96
Db 49 DLFEIEDREWIYEREFEEKEDVKEGVRVDLVFEGVDLSDVYLVNGVYLGSTEDMFIEYR 108
QY 97 -----NALKFLIQ-----GV-----TPLEYLAHRGFAHVGRHFTG-- 126
Db 109 FQVTNVLKEKNHLKVIKSPIRVPKTLQNGYGLGGPDPI-----RGYIRKAQYISYGD 163
QY 127 EGARIACQ-----MQSIDLRHYQTETHAMSTYNKFFNGFHH----- 163
Db 164 WGARIVTSGIWKVPYLEVYRARLQDSTAYLLEFGKDALVR-VNGFVHGEGLIVEVYVN 222
QY 164 -----SNQWFDRWYLSVPKSPFFEDAYSSGPFELTAVSFSEFYVLTNLLFV 210
Db 223 GEKIGEPVLEKNEKLEKLDGVFHLKDVKLWY-----PWNVGPYLYDFVFLKDL--- 272
QY 211 PFMSGAAYNGDMSTVTFGSAQSDSRHMTLGICIKELLEQDFDNVP-IQVRWIDKWF 269
Db 273 -----NGEI-----YREKK---ICLRRVRIVQEPDEGKTFIFEINCEKVFA 312
```

Search completed: September 26, 2000, 20:24:21  
Job time: 3645 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:28:54 ; Search time 92.32 Seconds  
(without alignments)  
387.526 Million cell updates/sec

Title: US-09-430-029-5  
Perfect score: 2873  
Sequence: 1 MDPTPLKKLGLKDRYAAMT.....FDGSEDQKNFAAWRGQATRN 516

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL12:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2372	82.6	536	Q9ZNP4	Q9znp4 comamonas t
2	2310.5	80.4	513	Q52572	Q52572 pseudomonas
3	2027	70.6	504	Q84961	Q84961 ralstonia s
4	2021	70.3	372	Q9ZNE3	Q9zne3 burkholderi
5	1854.5	64.5	516	Q52173	Q52173 pseudomonas
6	1849.5	64.4	516	Q52164	Q52164 pseudomonas
7	1799	62.6	511	Q32431	Q32431 acinetobact
8	1795	62.5	511	Q43981	Q43981 acinetobact
9	1542	53.7	377	Q92454	Q92454 pseudomonas
10	1502	52.3	381	Q92455	Q92455 pseudomonas
11	1493	52.0	381	Q92457	Q92457 pseudomonas
12	1354	47.1	327	Q92456	Q92456 pseudomonas
13	860	29.9	205	Q92458	Q92458 pseudomonas
14	844	29.4	204	Q92NE2	Q9zne2 comamonas s
15	795	27.7	215	Q92459	Q92459 pseudomonas
16	696.5	24.2	169	Q30590	Q30590 burkholderi
17	386	13.4	500	P95408	P95408 pseudomonas
18	377.5	13.1	498	O87798	O87798 pseudomonas
19	370	12.9	501	Q53027	Q53027 nocardia co

20	368.5	12.8	497	2	O87082	O87082 xanthobacte
21	354.5	12.3	501	2	O69178	O69178 alcaligenes
22	333.5	11.6	501	2	O51939	O51939 burkholderi
23	324	11.3	501	2	O07068	O07068 burkholderi
24	322	11.2	69	2	Q9ZNL2	Q9znl2 unidentified
25	320	11.1	69	2	Q9ZNL4	Q9znl4 unidentified
26	317	11.0	69	2	Q9ZNL7	Q9znl7 unidentified
27	279	9.7	69	2	Q9ZNL5	Q9znl5 unidentified
28	279	9.7	69	2	Q9ZNL1	Q9znl1 unidentified
29	278	9.7	69	2	Q9Z3A6	Q9z3a6 unidentified
30	271	9.4	69	2	Q9ZNL6	Q9znl6 unidentified
31	271	9.4	69	2	Q9ZNL3	Q9znl3 unidentified
32	263	9.2	69	2	Q9ZNL8	Q9znl8 unidentified
33	227.5	7.9	526	2	O06116	O06116 methylocyst
34	162	5.6	233	2	Q924R3	Q924r3 unidentified
35	124.5	4.3	174	2	O34217	O34217 uncultured
36	116	4.0	341	2	Q9ZET3	Q9zet3 xanthobacte
37	115	4.0	343	2	Q53025	Q53025 nocardia co
38	111.5	3.9	851	12	O85491	O85491 bovine leuk
39	111.5	3.9	1416	12	O92812	O92812 bovine leuk
40	110	3.8	1017	5	Q9XYV2	Q9xyv2 lymantria d
41	109	3.8	150	2	Q9XB08	Q9xbu8 uncultured
42	109	3.8	329	2	Q51943	Q51943 burkholderi
43	107.5	3.7	174	2	O34218	O34218 uncultured
44	106.5	3.7	1780	10	Q9ZT82	Q9zt82 arabidopsis
45	99.5	3.5	742	5	O61708	O61708 caenorhabdi

ALIGNMENTS

RESULT 1					
Q9ZNP4					
ID	Q9ZNP4	PRELIMINARY;	PRT;	536 AA.	
AC	Q9ZNP4;				
DT	01-MAY-1999 (TREMBLrel. 10, Created)				
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)				
DT	01-MAY-1999 (TREMBLrel. 10, Last annotation update)				
DE	PHENOL HYDROXYLASE COMPONENT.				
GN	APHN.				
OS	Comamonas testosteroni (Pseudomonas testosteroni).				
OC	Bacteria; Proteobacteria; beta subdivision; Comamonadaceae; Comamonas.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-TA441.				
RX	MEDLINE; 99018839.				
RA	ARAI H., AKAHIRA S., OHISHI T., MAEDA M., KUDO T.;				
RT	*Adaptation of Comamonas testosteroni TA441 to utilize phenol:				
RT	organization and regulation of the genes involved in phenol				
RT	degradation.*;				
RL	Microbiology 144:2895-2903(1998).				
DR	EMBL; AB006479; BAA34172.1; -.				
SQ	SEQUENCE 536 AA; 61860 MW; 29DB0C60 CRC32;				

Query Match	82.6%;	Score 2372;	DB 2;	Length 536;	
Best Local Similarity	81.4%;	Pred. No. 6.8e-197;			
Matches 416;	Conservative 45;	Mismatches 50;	Indels 0;	Gaps 0;	
QY	1	MDPTPLKKLGLKDRYAAMTRGLGWETTYQPMDKVPYDREYEGIKIHDWKVDVDFRLTM	60		
Db	1	MDAPVKKKLGKDRYATMTRGLGWETTSVQPMDKVPYDREYEGIKIHDWKVDVDFRLTM	60		
QY	61	DAYWYQGEKEKKLYAVDAFTQNNALFQVSDARYNALKLFQGVTPLEYLAHGRFAHV	120		
Db	61	DAYWYQGEKEKKLYAVFAFAQNNQGLGVTDAVYINALKLFQGVTPLEYLAHGRFAHA	120		
QY	121	GRHFTGEGARTACQMSIDELRHYOTETHAMSTYKFFNGFHHSNQWDFRVWVLSVPKSF	180		
Db	121	GRHFTGAGARAAQMSIDELRHFQTHAUSYKRYFNGMHSSSHWDFRVWVLSVPKSF	180		
QY	181	FDAYSSGPFPELTAVSFSEYVLNLLFVPPMGAAYNGDMSTVTFGFSQSDSRRMT	240		







DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DE 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE PHENOL HYDROXYLASE P3 PROTEIN (EC 1.14.13.7)  
 DE (PHENOL 2-MONOOXYGENASE P3 COMPONENT).  
 GN PHLD.  
 OS Pseudomonas putida.  
 OG Plasmid pPGH1.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
 OC Pseudomonas.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H;  
 RX MEDLINE; 95272534.  
 RA HERRMANN H., MUELLER C., SCHMIDT I., MAHNKE J., PETRUSCHKA L.,  
 RA HAHNKE K.;  
 RT "Localization and organization of phenol degradation genes of  
 Pseudomonas putida strain H.";  
 RL Mol. Genet. 247:240-246(1995).  
 CC -1- FUNCTION: CATABOLIZES PHENOL, AND SOME OF ITS METHYLATED  
 CC DERIVATIVES.  
 CC -1- CATALYTIC ACTIVITY: PHENOL + NADPH + O(2) = CATECHOL + NADP(+) +  
 CC H(2)O.  
 CC -1- COFACTOR: FAD FLAVOPROTEIN, AND REQUIRES PE(+2) FOR ACTIVITY (BY  
 CC SIMILARITY).  
 CC -1- PATHWAY: FIRST STEP OF PHENOL BIODEGRADATION.  
 CC -1- SUBUNIT: THE MULTICOMPONENT ENZYME PHENOL HYDROXYLASE IS FORMED BY  
 CC P0, P1, P2, P3, P4 AND P5 POLYPEPTIDES.  
 DR EMBL; X80765; CAA56743.1;  
 KW Aromatic hydrocarbons catabolism; Oxidoreductase; Monooxygenase;  
 KW Flavoprotein; FAD; Iron; Plasmid.  
 SQ SEQUENCE 516 AA; 60450 MW; 91455AA6 CRC32;

Query Match 64.4%; Score 1849.5; DB 2; Length 516;  
 Best Local Similarity 64.8%; Pred. No. 8.9e-152;  
 Matches 328; Conservative 65; Mismatches 106; Indels 9; Gaps 1;

QY 5 TLKKKLGKDRYAAMTRGLGWETTYQPMKVPYDREYEGIKHDWDKWDVDFRLTMDAYW 64  
 DB 3 TNKRLNLKDKYRLTRDLGWEPYQKEDVFPLEHREGIKITDWDKWDVDFRLTMDSYW 62  
 QY 65 KYQGEKKLYAVIDAFTONNAFLGVSDARYINALKFLQGVTPLEYLAHRGFHVGRHF 124  
 DB 63 KHQAEKKLYAIFAFAQNGHQNISDARYVNAKLFETGVSPEYQAFQGFGRVGRQF 122  
 QY 125 TGEARTACQMSIDELRHVQOTETHAMSTYNKFFNGFHSNQWDFRVWYLSVPKSFEDDA 184  
 DB 123 SGAGARVACQMAIDELRHVQOTVHAMSHYKHFDFGLHDFAHMYDRVWFLSVPKSFMDDA 182  
 QY 185 YSSGPEFLTAVSFSEYVLNLLFVPMGAAVNGDMSTVTFGFSAGSDSRHMTLGIE 244  
 DB 183 RTAGPEFLTAVSFSEYVLNLLFVPMGAAVNGDMSTVTFGFSAGSDSRHMTLGIE 242  
 QY 245 CIKELLEQDPNPVIVORWIDKWFWRGRLTLVAMMDYMQPKRVMSWRSWEMVABQN 304  
 DB 243 VIKEMLEQHDNPVIVORWIDKWFWRGRLTLVAMMDYMLPNKVMWSWSEAWGVPEQA 302  
 QY 305 GGALFKDLARYGIREPKGWQDACGCKDHIHQWSTYFGNAASFHTWVTEDEMGWLS 364  
 DB 303 GGALFKDLARYGIREPKGWQDACGCKDHIHQWSTYFGNAASFHTWVTEDEMGWLS 362  
 QY 365 AKYPDSDRYRPRFDHGWGEQARAGNRYMKTLPMLQCTQIPMLFTEPGNPTKIGARES 424  
 DB 363 EKYEDTDKYRPRFWRQEQAGERYNTLPHLCQVQCVPAIFTEPDDPTKLSRLSL 422  
 QY 425 NYLGNKHFCSHCKDIFDHEPQKYQAWLPVHOIHGNCFFPDADPGAEQDFDLAAVLD 484  
 DB 423 VHEGERVHFCSDGCCDIFKNEPVKYIQAWLPVHOIHGNCFFPDADPGAEQDFDLAAVLD 473  
 QY 485 YYATMGNDNLDFGSDQKNFAARQ 512  
 DB 474 YYHKSGVDNLEYLGSPEHQWLAALQG 501

RESULT 7  
 Q32431  
 ID Q32431 PRELIMINARY; PRT; 511 AA.  
 AC Q32431;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE DMS OXYGENASE COMPONENT.  
 DE DSOD.  
 GN Acinetobacter sp.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
 OC Moraxellaceae; Acinetobacter.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=20B;  
 RX MEDLINE; 98005684.  
 RA HORINOUCHE M., KASUGA K., NOJIRI H., YAMANE H., OMORI T.;  
 RT "Cloning and characterization of genes encoding an enzyme which  
 RT oxidizes dimethyl sulfide in Acinetobacter sp. strain 20B.";  
 RL FEMS Microbiol. Lett. 155:99-105(1997).  
 DR EMBL; D85083; BAA23333.1;  
 SQ SEQUENCE 511 AA; 60332 MW; CC36C8CE CRC32;

Query Match 62.6%; Score 1799; DB 2; Length 511;  
 Best Local Similarity 64.3%; Pred. No. 2e-147;  
 Matches 326; Conservative 59; Mismatches 112; Indels 10; Gaps 2;

QY 8 KKLGLKDRYAAMTRGLGWETTYQPMKVPYDREYEGIKHDWDKWDVDFRLTMDAYWYQ 67  
 DB 13 KKLNAKERYRLTRDLGWDFSYADKDAFPYEEFEGIKITDWSKWDVDFRLTMDYWKYQ 72  
 QY 68 GEKKEKLYAVIDAFTONNAFLGVSDARYINALKFLQGVTPLEYLAHRGFHVGRHFGE 127  
 DB 73 AEKKEKLYAIFAFAQNGHQNISDARYVNAKLFETGVSPEYQAFQGFGRVGRQFSGI 132  
 QY 128 GARTACQMSIDELRHVQOTETHAMSTYNKFFNGFHSNQWDFRVWYLSVPKSFEDAYSS 187  
 DB 133 GARTACQMSIDELRHVQOTETHAMSHYKHFDFGQDWAHMDRVWYLSVPKSFEDARSA 192  
 QY 188 GPFEFLTAVSFSEYVLNLLFVPMGAAVNGDMSTVTFGFSAGSDSRHMTLGIECIK 247  
 DB 193 GPFEFLTAVSFSEYVLNLLFVPMGAAVNGDMATVTFGFSAGSDSRHMTLGIEIVK 252  
 QY 248 FLEQDPNPVIVORWIDKWFWRGRLTLVAMMDYMQPKRVMSWRSWEMVABQN 307  
 DB 253 FLEQHDNPVIVORWIDKWFWRGRLTLVAMMDYMLPNKVMWSKDAWETTFEAGGA 312  
 QY 308 LFKDLARYGIREPKGWQDACGCKDHIHQWSTYFGNAASFHTWVTEDEMGWLSAKY 367  
 DB 313 LFKDLARYGIREPKGWQDACGCKDHIHQWSTYFGNAASFHTWVTEDEMGWLSAKY 372  
 QY 368 PDSFDRYRPRFDHGWGEQARAGNRYMKTLPMLQCTQIPMLFTEPGNPTKIGARESNY 426  
 DB 373 PDSFDRYRPRFDHGWGEQARAGNRYMKTLPMLQCTQIPMLFTEPGNPTKIGARESNY 432  
 QY 427 LGNKFHFCSDHCKDIFDHEPQKYQAWLPVHOIHGNCFFPDADPGAEQDFDLAAVLDY 486  
 DB 433 KDERVHTCSGCHDIFEREPEKYIQAWLPVHOIHGNCFFPDADPGAEQDFDLAAVLDY 483  
 QY 487 AVTMGRNDNLDFGSDQKNFAARQ 513  
 DB 484 NFNVGADNLDFGSDQKNFAARQ 510

RESULT 8  
 Q43981  
 ID Q43981 PRELIMINARY; PRT; 511 AA.  
 AC Q43981;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-NOV-1996 (TREMBLrel. 08, Last annotation update)

DE PHENOLHYDROXYLASE COMPONENT.  
 OS Acinetobacter calcoaceticus.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
 OC Moraxellaceae; Acinetobacter.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIB8250;  
 RX MEDLINE; 96154937.  
 RA EERT S., SCHIRMER F., HILLEN W.;  
 RT "Genetic organization, nucleotide sequence and regulation of  
 RT expression of genes encoding phenol hydroxylase and catechol 1,2-  
 RT dioxygenase in Acinetobacter calcoaceticus NCIB8250."  
 RL Mol. Microbiol. 18:13-20(1995).  
 DR EMBL; 236909; CAA85383.1; -.  
 SQ SEQUENCE 511 AA; 60316 MW; DA0F2BCE CRC32;

Query Match 62.5%; Score 1795; DB 2; Length 511;  
 Best Local Similarity 64.1%; Pred. No. 4.5e-147;  
 Matches 325; Conservative 60; Mismatches 112; Indels 10; Gaps 2;

QY 8 KTLGLKDRYAAATRGIGWETTYQPMQKVFVYDRYEGIKIHQDKWDPFRITMDAYWKYQ 67  
 Db 13 KLNKAKERYILTRDLWDYFADKADKADFEYEEFEGIKIDSKWEDPFRITMDYWKYQ 72  
 QY 68 GEKEKKLYAVIDAFTONNAFLGVSADARYINALKFLQGVTPLEYLAHGRFAHGRHFTGE 127  
 Db 73 AEKEKKLYAIFDAFAQNGGMNVSERYNAIKLFTAVTPLEYQAYQGYAHVGRQFSGI 132  
 QY 128 GARIACQMSIDELRHVQTHAMSTYKFNFGFHSNQWDRVWYLSVPKSFEDAYSS 187  
 Db 133 GARIASQMSIDELRHVQTHAMSHYKFNFGFQDWAHMDRVWYLSVPKSFEDARSA 192  
 QY 188 GPFEFTAVSFSEYVLTNLLFPFMSGAAYNGDMSTVTFGSAQSDSRHMTLIGICIK 247  
 Db 193 GPFEFTAVSFSEYVLTNLLFPFMSGAAYNGDMATVTFGSAQSDSRHMTLIGLIVK 252  
 QY 248 FILEQDPDNPVIVQWIDKFWRGYRLTLVAMMDYMQPKRYMSWRESWEMAYEQNGGA 307  
 Db 253 FILEQEDHNPVIVQWIDKFWRGYRLTLVAMMDYMLPKRYMSWRESWEMAYEQNGGA 312  
 QY 308 LFKDLARYGIREPKGWQDACEGKHSHQAWSTFYGFNAASAFHTWPTDEMGWLSAKY 367  
 Db 313 LFKDLARYGIRMPKYEVSISKEHSHQAWWIFNFGHAAGFHTWPTDEMDWLSKY 372  
 QY 368 POFSDRYRPRFDHMGEOQARAGNFYMKTLPLMLCQTCQIPMLETE-PGNPTKIGARESNY 426  
 Db 373 POFSDRYRPRWELARKMEAGKREYSAGLPOLCQVCQIPMTFTMDGDPFLSYRDSIY 432  
 QY 427 LGNKFHCSHDKDIFDHPQKYOQAWLPVHQHOGNCPFPDADGAGFDPALAAVLDYY 486  
 Db 433 KDERTCSGCHDIFEREPEKYOQAWLPVHQHOGNCPFPDLE-----SILRDYY 483  
 QY 487 AVTMGRDNLDFDGSDEQKNFAWRGQA 513  
 Db 484 NENVGADNLDEGSPDQQRKWKWNGNA 510

RESULT 9  
 Q92454 ID Q92454 PRELIMINARY; PRT; 377 AA.  
 AC Q92454;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-NOV-1999 (TREMblrel. 10, Last sequence update)  
 DE PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).  
 OS Pseudomonas putida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
 OC Pseudomonas.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=P-10;  
 RA FUTAMATA H., WATANABE K., SUDA F., HARAYAMA S.;

RT "Unique primary structure found in phenol hydroxylases Exhibiting high  
 RT affinity towards trichloroethylene.";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB016860; BAA74745.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 377 377  
 SQ SEQUENCE 377 AA; 44269 MW; 7B66BF49 CRC32;

Query Match 53.7%; Score 1542; DB 2; Length 377;  
 Best Local Similarity 71.6%; Pred. No. 2.1e-125;  
 Matches 270; Conservative 46; Mismatches 61; Indels 0; Gaps 0;

QY 63 YWKYQGEKEKKLYAVIDAFTONNAFLGVSADARYINALKFLQGVTPLEYLAHGRFAHGR 122  
 Db 1 YWKYQSEKERKLYAIDSFVQNGHNLNVDPRYLNALRLFTVTPLEYAARHYAHLGR 60  
 QY 123 HFTGEGARIAQMSIDELRHVQTHAMSTYKFNFGFHSNQWDRVWYLSVPKSEFE 182  
 Db 61 HFCGAGARVAAQMSIDELRHAQTLHTLSYKFNFGFGEWRHMDRVWYLSVPKSYFE 120  
 QY 183 DAYSSGPEFTAVSFSEYVLTNLLFPFMSGAAYNGDMSTVTFGSAQSDSRHMTLG 242  
 Db 121 DAMSAGPEFTAVSFSEYVLTNLLFPFMSGAAYNGDMATVTFGSAQSDSRHMTLG 180  
 QY 243 IECIKETLEQDPDNPVIVQWIDKFWRGYRLTLVAMMDYMQPKRYMSWRESWEMAYE 302  
 Db 181 LEVVKFCEQDPGNIPLLQKWLKFWRGYRLTLVAMMDYMLPKRYMSWRESWEMAYE 240  
 QY 303 QNGGALFKDLARYGIREPKGWQDACEGKHSHQAWSTFYGFNAASAFHTWPTDEMGW 362  
 Db 241 QAGGALFKDLARYGIRMPKYEVSISKEHSHQAWWIFNFGHAAGFHTWPTDEMDW 300  
 QY 363 LSAYKPSDFRYRPRFDHMGEOQARAGNFYMKTLPLMLCQTCQIPMLETEPGNPTKIGAR 422  
 Db 301 LAGKYPTFEYRPRLDHWOERQEGEFYFNATLPLMLCQTCQIPMVESEADDTQTQCYR 360  
 QY 423 ESNYLGKHFHCSHDK 439  
 Db 361 ESSYHGKHFHCSHDK 377

RESULT 10  
 Q92455 ID Q92455 PRELIMINARY; PRT; 381 AA.  
 AC Q92455;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-NOV-1999 (TREMblrel. 10, Last sequence update)  
 DE PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).  
 OS Pseudomonas putida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
 OC Pseudomonas.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=P-8;  
 RA FUTAMATA H., WATANABE K., SUDA F., HARAYAMA S.;

RT "Unique primary structure found in phenol hydroxylases Exhibiting high  
 RT affinity towards trichloroethylene.";  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB016859; BAA74744.1; -.  
 FT NON\_TER 1 1  
 FT NON\_TER 381 381  
 SQ SEQUENCE 381 AA; 44420 MW; 6B2A25F1 CRC32;

Query Match 52.3%; Score 1502; DB 2; Length 381;  
 Best Local Similarity 70.3%; Pred. No. 6.1e-122;  
 Matches 268; Conservative 44; Mismatches 69; Indels 0; Gaps 0;

QY 61 DAYWKYQGEKEKKLYAVIDAFTONNAFLGVSADARYINALKFLQGVTPLEYLAHGRFAH 120  
 Db 1 DTWKYQAEKEKKLYAIFDAFSCQNGHHTLSDARYNALKFLSGVTPLEYQAYQGFARV 60

QY 121 GRHFTGEGARIACOMSIDELRHVQTHAMSTYNKFNFGFHSNQWEDRWVWYLSVPKSF 180  
DB 61 GROFSGAGARVACQMAIDELRHVQTHAMSHYKNHFNGLHDFAHMHDWRWFLSVPKSF 120  
QY 181 FDAYSSTGPEFLTAVSFSEYVLTNLLFVPMGAAAYNGDMSTVTFGSAQSDSRHMT 240  
DB 121 FEDARTAGPEFLTAVSFSEYVLTNLLFVPMGAAAYNGDMATVTFGSAQSDSRHMT 180  
QY 241 LGIECIKFLLEQDPNPVIVQWIDKWFWRGYRLTLVAMMDYMQPKRVMWSRESWEMY 300  
DB 181 LGLEVIKFLLEQHDNPVIVQWIDKWFWRGYRLTLVAMMDYMLPNKVMWSSEANGVY 240  
QY 301 AEONGCALFKDLARYGIRPKHVEANIGKDHISHQAWSTFYGNAASAFHTWPTDEEM 360  
DB 241 FEEAGGALFKDLARYGIRPKHVEANIGKDHISHQAWSTFYGNAASAFHTWPTDEEM 300  
QY 361 GWLSAKYPSDFRYRPRDHWGEQARAGNRFYMKTLPMLCQTCQIPMLFTEPGNPTKIG 420  
DB 301 DMLSSKYPPTFKIYPRYEHWRALQEKGERFYNPTLPMLCQICQIPLFSGEPDDTTLS 360  
QY 421 ARESNYLGKHFHFCSDHCKDI 441  
DB 361 HRSVHEGERYHFCSQGCCDI 381

RESULT 11  
Q92457  
ID Q92457 PRELIMINARY: PRT; 381 AA.  
AC Q92457;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).  
OS Pseudomonas putida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
Pseudomonas.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=P-5;  
RA FUTAMATA H., WATANABE K., SUDA F., HARAYAMA S.;  
RT "Unique primary structure found in phenol hydroxylases Exhibiting high  
RT affinity towards trichloroethylene."  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB016857; BAA7473.1; -;  
DR EMBL; AB016854; BAA7473.1; -;  
FT NON\_TER 1  
FT NON\_TER 381  
SQ SEQUENCE 381 AA; 44525 MW; 5591FE88 CRC32;

Query Match 52.0%; Score 1493; DB 2; Length 381;  
Best Local Similarity 70.1%; Pred. No. 3.7e-121;  
Matches 267; Conservative 43; Mismatches 71; Indels 0; Gaps 0;

QY 61 DAYWKYGEKELIYVIDAFTONNAPLGVSDARYNALKLFLOGVTPLEYLAHRGPAHV 120  
DB 1 DTWKYQAEKKLIYAFDAFSONGHTTLDARYNALKLFLOGVTPLEYLAHRGPAHV 60  
QY 121 GRHFTGEGARIACOMSIDELRHVQTHAMSTYNKFNFGFHSNQWEDRWVWYLSVPKSF 180  
DB 61 GROFSGAGARVACQMAIDELRHVQTHAMSHYKNHFNGLHDFAHMHDWRWFLSVPKSF 120  
QY 181 FDAYSSTGPEFLTAVSFSEYVLTNLLFVPMGAAAYNGDMSTVTFGSAQSDSRHMT 240  
DB 121 FEDARTAGPEFLTAVSFSEYVLTNLLFVPMGAAAYNGDMATVTFGSAQSDSRHMT 180  
QY 241 LGIECIKFLLEQDPNPVIVQWIDKWFWRGYRLTLVAMMDYMQPKRVMWSRESWEMY 300  
DB 181 LGLEVIKFLLEQHDNPVIVQWIDKWFWRGYRLTLVAMMDYMLPNKVMWSSEANGVY 240  
QY 301 AEONGCALFKDLARYGIRPKHVEANIGKDHISHQAWSTFYGNAASAFHTWPTDEEM 360  
DB 241 FEEAGGALFKDLARYGIRPKHVEANIGKDHISHQAWSTFYGNAASAFHTWPTDEEM 300

DB 241 FEEAGGALFKDLARYGIRPKHVEANIGKDHISHQAWSTFYGNAASAFHTWPTDEEM 300  
QY 361 GWLSAKYPSDFRYRPRDHWGEQARAGNRFYMKTLPMLCQTCQIPMLFTEPGNPTKIG 420  
DB 301 DMLSSKYPPTFKIYPRYEHWRALQEKGERFYNPTLPMLCQICQIPLFSGEPDDTTLS 360  
QY 421 ARESNYLGKHFHFCSDHCKDI 441  
DB 361 HRSVHEGERYHFCSQGCCDI 381

RESULT 12  
Q92456  
ID Q92456 PRELIMINARY: PRT; 327 AA.  
AC Q92456;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).  
OS Pseudomonas putida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
Pseudomonas.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=P-6;  
RA FUTAMATA H., WATANABE K., SUDA F., HARAYAMA S.;  
RT "Unique primary structure found in phenol hydroxylases Exhibiting high  
RT affinity towards trichloroethylene."  
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB016858; BAA7474.1; -;  
FT NON\_TER 1  
FT NON\_TER 327  
SQ SEQUENCE 327 AA; 38388 MW; DFF0A7B0 CRC32;

Query Match 47.1%; Score 1354; DB 2; Length 327;  
Best Local Similarity 74.6%; Pred. No. 3e-109;  
Matches 244; Conservative 31; Mismatches 52; Indels 0; Gaps 0;

QY 65 KYQGEKELIYVIDAFTONNAPLGVSDARYNALKLFLOGVTPLEYLAHRGPAHVGRHF 124  
DB 1 KYQAEKKLIYAFDAFQNNHONISDARYNALKLFLOGVTPLEYLAHRGPAHVGRHF 60  
QY 125 TGGARIACOMSIDELRHVQTHAMSTYNKFNFGFHSNQWEDRWVWYLSVPKSFEDA 184  
DB 61 GGAGARVACQMAIDELRHVQTHAMSHYKNHFNGLHDFAHMHDWRWFLSVPKSFEDA 120  
QY 185 YSSGPPFELTAVSFSEYVLTNLLFVPMGAAAYNGDMSTVTFGSAQSDSRHMTLIE 244  
DB 121 RTAGPPFELTAVSFSEYVLTNLLFVPMGAAAYNGDMATVTFGSAQSDSRHMTLIE 180  
QY 245 CIKFLLEQDPNPVIVQWIDKWFWRGYRLTLVAMMDYMQPKRVMWSRESWEMYAFON 304  
DB 181 VIKFLLEQHDNPVIVQWIDKWFWRGYRLTLVAMMDYMLPNKVMWSAEWEVEQA 240  
QY 305 GGALFKDLARYGIRPKHVEANIGKDHISHQAWSTFYGNAASAFHTWPTDEMGWLS 364  
DB 241 GGALFKDLARYGIRPKHVEANIGKDHISHQAWSTFYGNAASAFHTWPTDEMGWLS 300  
QY 365 AKYPDSDRYRPRDHWGEQARAGN 391  
DB 301 AKYPDSDRYRPRDHWGEQARAGN 327

RESULT 13  
Q92458  
ID Q92458 PRELIMINARY: PRT; 205 AA.  
AC Q92458;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)  
DE PHENOL HYDROXYLASE ALPHA SUBUNIT (FRAGMENT).  
OS Pseudomonas putida.



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:11:44 ; Search time 75.19 Seconds

(without alignments)  
162.549 Million cell updates/sec

Title: US-09-430-029-5

Perfect score: 2873

Sequence: 1 MDTPTLKKLGLKDRYAAMT.....FDGSEDPQKNEAAWRGQATRN 516

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2409	83.8	519	1 W06802	Toluene ortho-mono
2	2027	70.6	504	1 W98973	Alcaligenes sp. pr
3	370	12.9	501	1 R66215	Nocardia corallina
4	370	12.9	501	1 R81471	Nocardia corallina
5	115	4.0	342	1 R66213	Nocardia corallina
6	109	3.8	343	1 R81469	Nocardia corallina
7	104	3.6	514	1 W12955	Bacillus sp. alpha
8	104	3.6	515	1 W31502	Bacillus sp. alpha
9	104	3.6	515	1 W48263	Bacillus sp. thalia
10	103	3.6	526	1 W59446	Arabidopsis thalia
11	102	3.6	572	1 P70579	Plasmid PRUB613 he
12	99	3.4	559	1 P70580	Plasmid PRUB616 he
13	98.5	3.4	684	1 W49873	Thermotoga maritim
14	94.5	3.3	1306	1 R04111	Human angiotensin
15	94	3.3	1965	1 W69165	Streptococcus pneu
16	93.5	3.3	738	1 Y01070	Human l(3)mht prot
17	93.5	3.3	772	1 Y01069	Human l(3)mht prot
18	92.5	3.2	843	1 R67760	Lys-aminopeptidase
19	92.5	3.2	1306	1 W68155	Human angiotensin
20	91.5	3.2	548	1 W39744	B. stearothermophi
21	91.5	3.2	548	1 Y01585	An alpha-amyrase (
22	91	3.2	485	1 W71249	Protein encoded by
23	91	3.2	626	1 Y01071	Mouse l(3)mht prot
24	90.5	3.2	1256	1 R27746	Muramidase release
25	90	3.1	217	1 W71251	Protein sequence o
26	90	3.1	549	1 P70338	Alpha-amyrase gene
27	89.5	3.1	962	1 R58701	Amino-peptidase H11
28	89.5	3.1	985	1 R42995	Glycosyltransferas
29	89.5	3.1	985	1 R42214	Aspergillus niger
30	88	3.1	514	1 R78269	Bacillus stearothe
31	88	3.1	515	1 R72449	Bacillus stearothe
32	88	3.1	515	1 W31406	Bacillus stearothe
33	88	3.1	713	1 R99797	Lysine decarboxyla

34	87	3.0	876	1 W21898	Alanyl-tRNA synthe
35	86.5	3.0	658	1 R23787	Heat-resistant pul
36	86.5	3.0	972	1 R51280	Helminth aminopept
37	86	3.0	771	1 W89589	Aspergillus oryzae
38	86	3.0	771	1 W97798	Aspergillus oryzae
39	85.5	3.0	826	1 Y01072	Rat l(3)mht protei
40	84.5	2.9	604	1 W13546	Human c-IAP2, Nucl
41	84.5	2.9	604	1 W19747	Human inhibitor of
42	84	2.9	518	1 R78680	Tobacco leaf antif
43	84	2.9	540	1 R78679	Tobacco leaf antif
44	83.5	2.9	529	1 P70376	Sequence of the al
45	83.5	2.9	1170	1 W88447	Yeast NPC1 protein

ALIGNMENTS

RESULT 1

W06802

ID W06802 standard; Protein; 519 AA.

AC W06802:

DE 29-JAN-1997 (first entry)

DE Toluene ortho-monoxygenase subunit tomA3.

KW tom; ptom; self-transmissible; constitutive; bioreactor; pollutant;

KW breakdown; trichloroethylene; TCE; degradation.

OS Pseudomonas cepacia strain PRL-23.

PN US5543317-A.

PD 06-AUG-1996.

PF 02-MAY-1991; 6947118.

PR 02-MAY-1991; US-694718.

PR 15-DEC-1993; US-167457.

PR 06-OCT-1994; US-319387.

PA (FRAN/) FRANCESCONI S C.

PA (SHLE/) SHIELDS M S.

PI Francesconi SC, Shields MS;

DR WPI; 96-370640/37.

DR N-PSDB; T44457.

PT Microorganisms transformed with P. cepacia PRL-23 Tom enzyme gene - are useful for degradation of chloro:aliphatic cpds. and aromatics

PS Claim 1; Column 29-32; 25pp; English.

CC The present sequence is that of toluene ortho-monoxygenase subunit

CC tomA3 encoded by T44457, isolated from Pseudomonas cepacia strain

CC PRL-23. The tom gene is present on a large self-transmissible plasmid

CC denoted ptom. The enzyme is capable of degrading trichloroethylene

CC (TCE), a hazardous pollutant. The ptom plasmid is transmissible and

CC expressable in other bacteria, thus many bacteria can be genetically

CC altered to constitutively degrade TCE, esp. in bioreactors or

CC TCE-contaminated environments. P. cepacia PRL-23 contg. ptom does not

CC need exogenous chemical inducers and is capable of functioning under a

CC diverse set of conditions. Also it does not require an inducer that is a

CC co-substrate for the enzyme required to break down TCE, in effect, it is

CC not subject to competitive inhibition.

SQ Sequence 519 AA;

Query Match 83.8%; Score 2409; DB 1; Length 519;

Best Local Similarity 81.8%; Pred. No. 5 9e-241;

Matches 419; Conservative 48; Mismatches 45; Indels 0; Gaps 0;

QY 1 MDTPTLKKLGLKDRYAAMT RGLGWETTYQPMDKVFPYDYEGIKIHWDKWDVDFPLTM 60

Db 1 MDTSVQKKLGLKDRYAAMT RGLGWETTYQPMDKVFPYDYEGIKIHWDKWDVDFPLTM 60

QY 61 DATWKYGEKEKLLYAVIDAFTQNNAPLGYSDARYINALKFLQGVTPLEYLAHGFARH 120

Db 61 DAYWKYGEKEKLLYAVIDAFAQNNQGLSDARYINALKFLQGVTPLEYLAHGFARH 120

QY 121 GRHETGEGARLACQMSIDELRYVOTETHAMSTYKFNFGFHHSNQWFDVWYLSVPKSF 180

Db 121 GRHETGEGARLACQMSIDELRYVOTETHAMSTYKFNFGFHHSNQWFDVWYLSVPKSF 180

QY 181 FEDAYSQGPFEFLTAVSFSEFYVLTNLLFVFFMSGAAYNGDMSTVTTFGSAQSDSRHMT 240

Db 181 FEDAYSQGPFEFLTAVSFSEFYVLTNLLFVFFMSGAAYNGDMSTVTTFGSAQSDSRHMT 240

```

Db 181 FEDATGGPEFELTAVSFSEYVLTNLLFVPFMSGAAAYNGDMSTVTGFGSAQSDSRHMT 240
QY 241 LGTECKFLEQDPNDVPIVQWRIDKWFWRGYRLLITIVAMMDYMPKRYMSWRESWEY 300
Db 241 LGTECKFLEQDPNDVPIVQWRIDKWFWRGYRLLITIVAMMDYMPKRYMSWRESWEY 300
QY 301 AEONGGALKOLARYGIREPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTWPTDEDM 360
Db 301 VEONGGALKOLARYGIREPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTWPTDEDM 360
QY 361 GWLSAKYPSDFRYPREDHWGEQARAGNRFYMKTLPMLCOTQCPMLFTEPGNPTKIG 420
Db 361 GWLSAKYPSDFRYPREDHWGEQARAGNRFYMKTLPMLCOTQCPMLFTEPGNPTKIG 420
QY 421 ARESNVLGNKHFCSCHDKDIEDHEPQKYQAWLPVHQIHQGNCFPPDADPGAEGEDPLA 480
Db 421 YRESAYLGDYHFCSDHCKEIDFNEPEKVSQWLPVQVYQGNCFPPDADPGAEGEDPLM 480
QY 481 AVLDDYAVTMGRNLDLFDGSEQKFNFAWRG 512
Db 481 ALLDYINLVGRNDFEGSEQKFNFAWRGE 512

```

## RESULT 2

```

W98973
ID W98973 standard; Protein: 504 AA.
AC W98973;
DT 10-MAY-1999 (first entry)
DE Alcaligenes sp. protein PoxD.
KW Alcaligenes; PoxR; PoxA; PoxB; PoxC; PoxD; PoxE; PoxF; PoxG; PoxH;
KW PoxI; aromatic; oxidative; petroleum purification; chemical industry;
KW drug.
OS Alcaligenes sp.
PN J11042088-A.
PD 16-FEB-1999.
PF 25-JUL-1997; 200625.
PR 25-JUL-1997; JP-200625.
PA (TOFU ) TONEN CORP.
DR WPI: 99-197820/17.
DR N-PSDB: X18867.
PT New encoding an aromatic cpd. oxidative decompsn. enzyme - useful in
PT the fields of petroleum purification, chemical and drug industries
PS Claim 1: Page 7-19; 35pp; Japanese.
CC The present sequence encodes Alcaligenes sp. protein PoxD. The present
CC invention also describes PoxA, PoxB, PoxC, PoxE, PoxF, PoxG, PoxH
CC and PoxI. The proteins are useful in the fields of petroleum
CC purification, chemical industry and drug industry related to the
CC synthesis, conversion and decomposition of aromatic compounds.
SQ Sequence 504 AA;

```

```

Query Match 70.6%; Score 2027; DB 1; Length 504;
Best Local Similarity 70.3%; Pred. No. 2e-201;
Matches 355; Conservative 62; Mismatches 78; Indels 10; Gaps 1;

QY 7 KKKLGLKDRVAAWGLGWETTYQPMKVPFYDREYEGIKIHDWDKWDVDFRLTMDAYWKY 66
Db 4 RKKLNLKRYATMRDUGWETTYEPMKVPFYDREYEGIKIHDWDKWDVDFRLTMDAYWKY 63
QY 67 QGEKELIAYIDAFTQNNAPLGVSDARYINALKFLQGVTPLEYLAHGFVHGRHFTG 126
Db 64 QSEKERLIAYIDSFVQNGHNLVSDARYINALKFLGTGTPTLEYLAHGFVHGRHFTG 123
QY 127 EGARIACOMSIDELRHQYQETHAMSTYKFFNGFHHSNQWDFRVVYLSVPKSFEDAYS 186
Db 124 AGARVAAQMSIDELRHAQQLHLSYVNYKFFGFGWRHMDRVVYLSVPKSFEDAYS 183
QY 187 SGPEFLTAVSFSEYVLTNLLFVPFMSGAAAYNGDMSTVTGFGSAQSDSRHMTLGIECI 246
Db 184 AGPFELTAVSFSEYVLTNLLFVPFMSGAAAYNGDMSTVTGFGSAQSDSRHMTLGIEV 243
QY 247 KFLLEQDPNDVPIVQWRIDKWFWRGYRLLITIVAMMDYMPKRYMSWRESWEYAEONGG 306
Db 247 KFLLEQDPNDVPIVQWRIDKWFWRGYRLLITIVAMMDYMPKRYMSWRESWEYAEONGG 306

```

```

Db 244 KFLCEQDPNDLPIQLQKWLDKWFWGRFLLTLVGMMDYMLPRRYMSWAEAEWTFEQAGG 303
QY 307 ALFKDLARYGIREPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTWPTDEMGWLSAK 366
Db 304 ALFKDLARYGIREPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTWPTDEMGWLSAK 363
QY 367 YPDSFDRYPREDHWGEQARAGNRFYMKTLPMLCOTQCPMLFTEPGNPTKIGARESNI 426
Db 364 YPOTFARYYPRLDYWOERQOAGERYNGHLPMLCOTQCPMLFTEPGNPTKIGARESNI 423
QY 427 LGNKFHFCSDHCKDIEDHEPQKYQAWLPVHQIHQGNCFPPDADPGAEGEDPLAALVDY 486
Db 424 HGMKFHFCSDGCKDIEDHEPEKVAQAWLPVHQIYQGNCFPPDADPGAEGEDPLAALVDY 486
QY 487 AVTMGRNLDLFDGSEQKFNFAWRG 511
Db 474 RLNLGADNLDFEGSEQKFNFAWRG 498

```

## RESULT 3

```

R66215
ID R66215 standard; Protein: 501 AA.
AC R66215;
DT 04-AUG-1995 (first entry)
DE Nocardia corallina alkene monooxygenase subunit-2.
KW alkene monooxygenase; subunit-2; epoxidation; amoC.
OS Nocardia corallina.
PN J06292571-A.
PD 21-OCT-1994.
PF 06-APR-1993; 105171.
PR 06-APR-1993; JP-105171.
PA (NIHA ) JAPAN ENERGY CORP.
DR WPI: 95-009069/02.
DR N-PSDB: Q79569.
PT Alkene monooxygenase and corresp. gene - useful for the
PT epoxidation of an alkene
PS Claim 1: Page 3-4; 30pp; Japanese.
CC E.coli transformed with the DNA sequence Q79569 are able to
CC catalyze the epoxidation of alkenes. The DNA is derived from
CC Nocardia corallina and comprises 4 open reading frames. ORFs amoA
CC and amoC encode subunits 1 and 2 of the alkene monooxygenase enzyme
CC and ORF amoD encodes a reductase capable of transferring electrons
CC from NADH coenzyme to a monooxygenase.
SQ Sequence 501 AA;

```

```

Query Match 12.9%; Score 370; DB 1; Length 501;
Best Local Similarity 25.6%; Pred. No. 7.9e-30;
Matches 124; Conservative 68; Mismatches 190; Indels 102; Gaps 21;

QY 37 PYDRYEGIKIHDWD-----KWVDPFLTMDAYWKYOGKEKKLYAVI 78
Db 5 PTQLHEKSKSDWDFTSVERRPKFKETKMKPKGKDPFVLIRDYMKMAEKDTRHGFL 64
QY 79 DAFTQNNAPLGVSDARYINALKFLQGVTPLEYLAHGFVHGRHFTGEGARIACOMQSI 138
Db 65 DGAVTREATRI-EPRFAEMKIMVQPTNAEYQAVACGCMILISAVENQELRQGYAAQML 123
QY 139 DELRHQYQETHAMSTYKFFNGFHHSNQWDFRVVYLSVPKSF----- 180
Db 124 DEVRHAQLE---MTLRN-----YYAKHWCD-----PSGFDIGORGLYOHAPGLVSIG 167
QY 181 -FEDAYSSGPFEFLTAVSFSEYVLTNLLFVPFMSGAAAYNGDMSTVTGFGSAQSDSRHMT 239
Db 168 EFQIFNTGDPDLVDITIDUNIVAEATFTNLLVATFPQAVANGDNAMASVFLTQSDARHM 227
QY 240 TLGIECKFLEQDPNDVPIVQWRIDKWFWRGYRLL-TLVAMMDYMPKRYMSWRESWE 298
Db 228 ANGYGSVMALLENE-DNLPLLNQSLDRHFWRAHRAKLDNAVGCWSEYGARCKPWSYKAQWE 286
QY 299 MYAEQN--GGALFKDLARYGIREPKGWQDACEG-----KDHISHQAWSTFYGFNAAS 348
Db 287 EWVVDVFGVGYIDR-LSEFGVQAP-----ACLGAAADEVKWSHHTLGOVL SAVPLN--- 337

```



```

QY 349 AFHTWVPTEDMG-----WLSAKYPDSFDRYRPRFDHW-GEQA---RAGNRFYKTLPM 399
D 338 ---FW---RSDAMGPADEWFENHYPGSAAYQ-----ACUGAAAEVKNSHHTLGQVLSAVWPLN--- 388
QY 400 L---CQTCQIPMLTEFCNPTKIGARESNYLGNKFHFCSDHCKDIFDHEPKY---VQAW 453
D 389 LPPMCQVCQVPCVMP---RLDMNAARILIEFGQIALCSEPCQRLFTNWPEAYRHRQYW 445
QY 454 LPVH 457
D 446 ARYH 449

RESULT 4
ID R81471 standard; Protein; 501 AA.
AC R81471;
DE 07-AUG-1996 (first entry)
KW Nocardia corallina alkene mono-oxygenase gene product, amoC.
KW Alkene mono-oxygenase; indole; indigo production; biosynthesis;
OS Nocardia corallina B-276.
FH Key Location/Qualifiers
FT misc_difference 268
FT /note= "Tyr residue given in the specification
FT corresponding to a TCG codon in T17418"
PN J08023988-A.
PD 30-JAN-1996.
PF 08-JUL-1994; 179688.
PR 08-JUL-1994; JP-179688.
PA (NIHA ) JAPAN ENERGY CORP.
DR WPI: 96-133426/14.
DR N-PSDB; T17418.
PT Prep'n. of indigo by a microbiological method - by culturing a
PT microbe having alkene monooxygenase activity to oxidise indole to
PT indigo
PS Claim 5; Page 7-8; 11pp; Japanese.
CC R81469-R81472 are protein products of the Nocardia corallina strain
CC B-276 alkene mono-oxygenase gene. The gene encodes 4 protein products
CC amoA, amoB, amoC and amoD derived from the 3 different reading frames
CC of the operon. The gene is useful for the production of indigo via
CC oxidation of indole. Nocardia corallina can be cultured in a medium
CC contg. indole and will readily oxidise the indole yielding indigo into
CC the culture medium. E. coli may also be transformed with the alkene
CC mono-oxygenase gene and used as above to efficiently produce indigo
CC by microbial oxidation.
SQ Sequence 501 AA;

Query Match 12.9%; Score 370; DB 1; Length 501;
Best Local Similarity 25.6%; Pred. No. 7.9e-30;
Matches 124; Conservative 68; Mismatches 190; Indels 102; Gaps 21;

QY 37 PYDRYEGIKTHDWD-----KWVDPRLTMDAYWKYQCEKELKLYAVI 78
D 5 PTOLHEKSKSVDDWTSVERRPKEETKYKPKKGDPRVILIRDMKMEAKDRTTGGF 64
QY 79 DAFTONNAFLGVSDARYINALKFLQGVTPLEYLAHGRFAHVRHFTGEGARIACQMOSI 138
D 65 DGAVRTREATRI-EPFAEAMKIMVOLTNAEYQAVAGCGMISAVENQELRQYAAQML 123
QY 139 DELRHYOTETHAMSTYKFFNGPHSHSNQWDFRWYLSVPKSF----- 180
D 124 DEVRHAQLE---MTLRN-----YIAKHWC-----PSGFDIGQRGLYQHAGLYSIG 167
QY 181 -FEDASSGPFEEITAVSFSEFYVLLNLFVPMFMSGAAYNGDMSTVTFGSAQSDSRHM 239
D 168 EFQHEHTGDPDLVIDLNIIVAEATAFTNLLIVATPQAVANGDNAMASVFLSIQSDEARM 227
QY 240 TLGIEICKFLEQDPNPVIVQKWDKFWGRYLL-TIVAMMDYMQPKRVMWSRESWE 298
D 228 ANGYGSMALLENE-DNLPLLSQDRHFWRAHKALDNVAGWCSEYGARPKRPSYKAQWE 286

```

```

QY 299 MYAEQN---GGALFKDLARYGIREPKGWODACEG-----KDHISHQAWSTFFYGFNAAS 348
D 287 EWWVDDDFVGGYIDR-LSEFGVQAP-----ACUGAAAEVKNSHHTLGQVLSAVWPLN--- 337
QY 349 AFHTWVPTEDMG-----WLSAKYPDSFDRYRPRFDHW-GEQA---RAGNRFYKTLPM 399
D 338 ---FW---RSDAMGPADEWFENHYPGSAAYQ-----ACUGAAAEVKNSHHTLGQVLSAVWPLN--- 388
QY 400 L---CQTCQIPMLTEFCNPTKIGARESNYLGNKFHFCSDHCKDIEDHEPKY---VQAW 453
D 389 LPPMCQVCQVPCVMP---RLDMNAARILIEFGQIALCSEPCQRLFTNWPEAYRHRQYW 445
QY 454 LPVH 457
D 446 ARYH 449

RESULT 5
ID R66213 standard; Protein; 342 AA.
AC R66213;
DE 04-AUG-1995 (first entry)
KW Nocardia corallina alkene mono-oxygenase subunit-1.
KW alkene mono-oxygenase; subunit-1; epoxidation; amoA gene.
OS Nocardia corallina.
PN J06292571-A.
PD 21-OCT-1994.
PF 06-APR-1993; 105171.
PR 06-APR-1993; JP-105171.
PA (NIHA ) JAPAN ENERGY CORP.
DR WPI: 95-005069/02.
DR N-PSDB; Q79569.
PT Alkene monooxygenase and corresp. gene - useful for the
PT epoxidation of an alkene
PS Claim 1; Page 2; 30pp; Japanese.
CC E.coli transformed with the DNA sequence Q79569 are able to
CC catalyse the epoxidation of alkenes. The DNA is derived from
CC Nocardia corallina and comprises 4 open reading frames, ORFs amoA
CC and amoC encode subunits 1 and 2 of the alkene monooxygenase enzyme
CC and ORF amod encodes a reductase capable of transferring electrons
CC from NADH coenzyme to a monooxygenase.
SQ Sequence 342 AA;

Query Match 4.0%; Score 115; DB 1; Length 342;
Best Local Similarity 19.8%; Pred. No. 0.0011;
Matches 49; Conservative 39; Mismatches 117; Indels 42; Gaps 6;

QY 44 IKIHWDKWKVDPFRLTMDAYWKYQCEKELKLYAVIDAFTQNNAFVGVSDARYINALKFL 103
D 58 VRTSEWSAYRDPHQLWQRPYVSTCNQDQQLARLVPVLTGMSA----- 110
QY 104 QGVTPLE-----EYLAHGRFAHVRHFTGEGARIACQMOSIDELRHYOTE 147
D 111 -AUIPWISQILARSYAAMPFVEYGLFSLAYAVRQAMSDTVQFVSWFQAVDRMELQDI 169
QY 148 THAMSTYKNF--FNGFHHSNQWDFRWYLSVPKSFEDFAYSSGPFEEITAVSFSEFYVLT 205
D 170 VVHLDHLQESPEFSDAGAREAMSDSTLVPREVIERIAASQDWWVEIIVAGTLVFEPLVG 229
QY 206 NLLFVP-FMSCAAAYNGDMSTVTFGSAQSDSRHMTIGIECKFLEQDP-----DNVPIV 260
D 230 HLAKAEFLSRAPMGPDGTTTAVLASALLDSGRHLE-SVQALVRLVCQDPVHGQDQNAV 288
QY 261 QRWIDKW 267
D 289 RWLEEW 295

RESULT 6
ID R81469 standard; Protein; 343 AA.

```

AC R81469;  
DT 07-AUG-1996 (first entry)  
DE Nocardia corallina alkene mono-oxygenase gene product, amoA.  
KW Alkene mono-oxygenase; indole; indigo production; biosynthesis;  
KW microbial oxidation; dye.  
OS Nocardia corallina B-276.  
PN J08023988-A.  
PD 30-JAN-1996.  
PF 08-JUL-1994; 179688.  
PR 08-JUL-1994; JP-179688.  
PA (NIHA ) JAPAN ENERGY CORP.  
DR WPI: 96-133426/14.  
DR N-PSDB; T17418.  
PT Prep. of indigo by a microbiological method - by culturing a  
PT microbe having alkene monoxygenase activity to oxidise indole to  
PT indigo.  
PS Claim 5: Page 6-7; l1pp; Japanese.  
CC R81469-R81472 are protein products of the Nocardia corallina strain  
CC B-276 alkene mono-oxygenase gene. The gene encodes 4 protein products  
CC amoA, amoB, amoC and amoD derived from the 3 different reading frames  
CC of the operon. The gene is useful for the production of indigo via  
CC oxidation of indole. Nocardia corallina can be cultured in a medium  
CC contg. indole and will readily oxidise the indole yielding indigo into  
CC the culture medium. E. coli may also be transformed with the alkene  
CC mono-oxygenase gene and used as above to efficiently produce indigo  
CC by microbial oxidation.  
SQ Sequence 343 AA;

Query Match 3.8%; Score 109; DB 1; Length 343;  
Best Local Similarity 19.4%; Pred. No. 0.0047;  
Matches 48; Conservative 39; Mismatches 118; Indels 42; Gaps 6;

QY 44 IKIHWDKWPFFRLTMDAYWYQGEKELIYAVIDAFTONNAFLGVSDDARYINALKLFL 103  
DB 69 VRTSESNAYRPHQLWQRPVYSTCNCQQAALRPVLTMGSA----- 111  
QY 104 QGVTPPL-----EYLAHGFAGVGRHFTGEGARTACQMSIDELRHVQTE 147  
DB 112 -AUIPIWSQILARSYAAPVEYGLFSLAYAVRQAMSDTVQSFVQVDRMRLLODI 170  
QY 148 THAMSTYNKF--FNGFHHSNWFDRVWYLSVPKFFEDAYSGPPEFTAVSFSEYVLT 205  
DB 171 VHLDLHGESPEFSDAGAREAWMSDSTLPIEVIERIAASQDWVELLVAGTLVEPLVG 230  
QY 206 NLLFVP-FWGAAYNGDMSTVTFGFSQAQSDSRHMTLGEICIKFLLEQDP-----DNVPIV 260  
DB 231 HLAKELEFSRRAPMGDGTTPPAVLASALLDSGRHLE-SVQALVRLVCDPVRHGDQNAIV 289  
QY 261 QRWIDKW 267  
DB 290 RWIEEW 296

RESULT 7  
W12955  
ID W12955 standard; protein; 514 AA.  
AC W12955;  
DT 07-APR-1997 (first entry)  
DE Alpha-amylase.  
KW Alpha-amylase; detergent; thermal stability; oxidation stability; enzyme;  
KW calcium ion dependency; alpha-amylolytic activity; washing composition;  
KW textile desizing; papermaking; beer-making; ethanol production;  
KW sweetener.  
OS Bacillus stearothermophilus.  
PN W09623873-A1.  
PD 08-AUG-1996.  
PF 05-FEB-1996; PF DK00056.  
PR 03-FEB-1995; DK-000126.  
PR 29-MAR-1995; DK-000336.  
PR 29-SEP-1995; DK-001097.  
PR 06-OCT-1995; DK-001121.  
PA (NOVO ) NOVO-NORDISK AS.

PI Bisgard-frantzen H, Borchert T, Svendsen A;  
DR WPI: 96-371423/37.  
PT Alpha-amylase variants - with improved thermal and oxidation  
PT stability and reduced calcium ion dependency  
PS disclosure; Page 86-88; l1lpp; English.  
CC W12955, W12956, and R81836 represent the parent alpha-amylases  
CC used to create the variants of the invention (such as W12098-W12131).  
CC This sequence represents the alpha-amylase from Bacillus  
CC stearothermophilus. The variants of the invention were created using  
CC site directed, or random, mutagenesis of the DNA sequences encoding these  
CC parent alpha-amylases. The variants of the invention can have improved  
CC thermal stability (such as at temperatures in the range of 40-70 degrees  
CC Celsius), and/or oxidation stability, and/or reduced calcium ion  
CC dependency. The variants can also have increased alpha-amylolytic  
CC activity (especially at pH values in the range of 8.5-10.5), and  
CC improved binding of a particular substrate. The variant alpha-amylases  
CC also possess improved specificity to a particular substrate, and/or  
CC improved specificity with respect to cleavage of substrate, and for textile  
CC can be used in detergent and washing compositions, and for papermaking and  
CC beer-making processes. The variants of the invention can also be used in  
CC the production of sweeteners and ethanol from starch.  
SQ Sequence 514 AA;

Query Match 3.6%; Score 104; DB 1; Length 514;  
Best Local Similarity 18.3%; Pred. No. 0.029;  
Matches 82; Conservative 43; Mismatches 125; Indels 198; Gaps 24;

QY 21 RGLG--WETTYQPMKVPFYDYREGI-----KIHWDKW-----VDPFRUTMDAY 63  
DB 179 RGIGKANDWEVDTEGNGYDILMYADLMDHDEVTTELKSWKQWYVNTNIDGRLDAVKH 238  
QY 64 WKYQGEKEKLYAVIDAFTONNAFL--GVSDARYINALKLFLQGVTPLEYLAHGFAGV 121  
DB 239 IKF-----SFPDWLSDVRSQTGKPLFTVG-----EYWSY----- 268  
QY 122 RHFTGEGARIACOMSIDELRHVQTEHMTSTYKFFNGFHHSNQWFEDEVWLSVPKSF 181  
DB 269 -----DINKLHNYIMKTN--GTMSLFDAPLHN-----KPY 296  
QY 182 EDAYSGPPEFTAVSFSEYVLTNLLFPVFMGAAINGDMSTVTFGFSQAQSDSRHMTL 241  
DB 297 TASKSGGTFDMKT-----LMTNTLM-----KQDTL----- 322  
QY 242 GIECIKFLLEQDPDNVPIVQRWIDKW-----WRGYRL----- 275  
DB 323 ----AVTFVNDHTEPGQALQSWDPWFKPLAYAFILTRQEGYPCVFGYGYGIPQYNIPS 379  
QY 276 -----TLVA-----MMMDYMQPKRVMSW-RE-----SWEM 299  
DB 380 LKSKIDPILLIARRDYAYGTQHDYLDHSDIIGTWREGVTEKPGSGLAALITDGGGSKWY 439  
QY 300 YAEQNGGALFKDLARYGIREPKGWQDACEKGKHISHQAMSTFYGFNAASAFHTWVP---T 356  
DB 440 VQKHAGKVFYDLT--GNR-----SDVT-----INSDGWGEF-KVNGGSV-SWVPRKIT 486  
QY 357 EDEMGMWLSAKYP--DSFDRYRPRFDHW 382  
DB 487 VSTIANSITRPTWDEVRWTEPLVAV 514

RESULT 8  
W31502  
ID W31502 standard; protein; 515 AA.  
AC W31502;  
DT 08-APR-1998 (first entry)  
DE Bacillus sp. alpha amylase.  
KW Alpha amylase; hard surface cleaning; dishwashing; laundry.  
OS Bacillus sp.  
PN W09732961-A2.  
PD 12-SEP-1997.  
PF 04-MAR-1997; U03635.



CC stimulate a defence response in transgenic plants conferring increased  
 CC pathogen resistance, especially resistance to powdery mildew or rust.  
 CC The product can also be used for identifying compounds able to  
 CC stimulate a defence response in a plant by interaction with  
 CC encoded polypeptide.  
 SQ Sequence 526 AA;

Query Match 3.6%; Score 103; DB 1; Length 526;  
 Best Local Similarity 22.0%; Pred. No. 0.038;  
 Matches 71; Conservative 43; Mismatches 130; Indels 78; Gaps 17;

QY 45 KIHWDKWDVDFRLTMDAYWKYQGS---KEKKLYAVIDAFTONNAFLVGSVDARYI--- 96  
 DB 184 RIHOWKWKEDST---ADEKFPDPTALRKRRVTHVHNHAFIKH-FLGIGKDSVILGWT 237  
 QY 97 -NALKLEQLGQVTPLEYLAHR-GFAHVGRRHFTGEGARIACOMQSIDELRHYQTETHAMSTY 154  
 DB 238 QSLKQFVDSYTKSDYVTLRLGF--IMTHCKG-NPKLNFHKYMMRALEDDEKQVVGISWY 294  
 QY 155 NKFF-----NGFHHSNQWDRWYLSVPKSEFFEDAYSQGPPEFLTAVSFSEYVLT 205  
 DB 295 LWIFWVIFLLNVNGWH-----TYFWIAFTI--PPALLAVGKLEHVIA 336  
 QY 206 NLLFVPMFMSGAAYNGDM---STVTFGFSQAQSDSRHMTLGTGICIKFLLEODPDNPVIVQ 261  
 DB 337 QLAHEVAEKHVAIGDLVVKFSDEHFWFSKQ-----IVLYLHLFILFQNAFEIAFF- 388  
 QY 262 RWIDKFWFRGKRLTLVAMMDYMOFK-----RVMSWRSEWEMYA--EQNGG---AL 308  
 DB 389 ---FWIWTYGFDSICMGQVRYIVPRLVIGVFIQVLCYSISLPLIAYIVSQMSGSFKKAI 444  
 QY 309 FKDLARYGIREPKGWODACEGK 330  
 DB 445 FEENVQVGL---VGWAQKVQK 463

## RESULT 11

ID P70579 standard; Protein; 572 AA.  
 AC P70579;  
 DT 07-MAY-1991 (first entry)  
 DE plasmid pTUB613 heat resistant alpha-amylase product.  
 KW Thermostable.  
 OS Bacillus stearo-thermophilus.  
 PN J62104580-A.  
 PD 15-MAY-1987.  
 PF 30-OCT-1985; 241302.  
 PR 30-OCT-1985; JP-241302.  
 PA (HIGE-) HIGETA SHOYU KK.  
 DR WPI: 87-173697/25.  
 DR N-PSDB; N70920.  
 PT DNA expressing highly-thermostable enzyme - formed by adding one  
 PT more cysteine codon in structural gene  
 PS Disclosure; Fig 6; 12pp; Japanese.  
 CC The enzyme is modified from the wild type by the inclusion of a  
 CC cysteine residue, increasing the thermostability of the alpha-  
 CC amylase.  
 SQ Sequence 572 AA;

Query Match 3.6%; Score 102; DB 1; Length 572;  
 Best Local Similarity 18.1%; Pred. No. 0.054;  
 Matches 81; Conservative 44; Mismatches 125; Indels 198; Gaps 24;

QY 21 RGLG--WETTYQPMKVFYDRIYEGT-----KIHWDKW-----VDPFRLTMDAY 63  
 DB 236 RGIGKAWDEVDTEGNYDILYADLMDHPVETLKSQGWKYVNTNIDGFRDLDAVKH 295  
 QY 64 WKYQGEKKLYAVIDAFTONNAFL--GYSDARYINALKFLQGVTPLEYLAHRGAHV 121  
 DB 296 IKF-----SFFPDWLSNVRSQTKPLFTVG-----EWSY----- 325

QY 122 RHFTGEGARIACOMQSIDELRHYQTETHAMSTYKFFNGFHHSNQWDRVWYLSVPKSEFF 181  
 DB 326 -----DINKLHNYNMTN--GTMSLFDAPLHN-----KFY 353  
 QY 182 EDAYSSGPFELTAVSFSEYVLTNLLFVPMGSAAYNGDMSTVTFGFSQAQSDSRHMTL 241  
 DB 354 TASKSGGTEDMRT-----LMTNTLM-----KDQPTL----- 379  
 QY 242 GIECIKFLLEQDPDNPVIVQWIDKWF-----WRGYRLL----- 275  
 DB 380 ---AVTFVDNHDTEPGALQSWVDWPKFLAYAFILTRQEGIPCVCYGYDYGIPQYNIPS 436  
 QY 276 -----TLVA-----MMMDYMQPKVMSW-RE-----SWEM 299  
 DB 437 LKSKIDPLLIARRDYAYGTQHDYLDHSDILGWTREGVTEKPGSLAALITDGPGRKWMY 496  
 QY 300 YAEQNGGALFKDLARYGIREPKGWODACEGKDHISHQAWSTYGFNAASAFHTWVP---T 356  
 DB 497 VGKQHAGKVFYDLT--GNR-----SDTVT-----INSDGMGEF-KVNGGSV-SWVVPKRKT 543  
 QY 357 EDEMGLWSAKYP--DSFDYRYRPRFDHW 382  
 DB 544 VSTIAWSITRPTWDETFVRWTEPRIVAW 571

## RESULT 12

ID P70580 standard; Protein; 559 AA.  
 AC P70580;  
 DT 07-MAY-1991 (first entry)  
 DE plasmid pTUB616 heat resistant alpha-amylase product.  
 KW Thermostable.  
 OS Bacillus stearo-thermophilus.  
 PN J62104580-A.  
 PD 15-MAY-1987.  
 PF 30-OCT-1985; 241302.  
 PR 30-OCT-1985; JP-241302.  
 PA (HIGE-) HIGETA SHOYU KK.  
 DR WPI: 87-173697/25.  
 DR N-PSDB; N70926.  
 PT DNA expressing highly-thermostable enzyme - formed by adding one  
 PT more cysteine codon in structural gene  
 PS Disclosure; Fig 7; 12pp; Japanese.  
 CC The enzyme is modified from the wild type by the inclusion of a  
 CC cysteine residue, increasing the thermostability of the alpha-  
 CC amylase.  
 SQ Sequence 559 AA;

Query Match 3.4%; Score 99; DB 1; Length 559;  
 Best Local Similarity 18.1%; Pred. No. 0.11;  
 Matches 81; Conservative 44; Mismatches 125; Indels 198; Gaps 24;

QY 21 RGLG--WETTYQPMKVFYDRIYEGT-----KIHWDKW-----VDPFRLTMDAY 63  
 DB 223 RGIGKAWDEVDTEGNYDILYADLMDHPVETLKSQGWKYVNTNIDGFRDLDAVKH 282  
 QY 64 WKYQGEKKLYAVIDAFTONNAFL--GYSDARYINALKFLQGVTPLEYLAHRGAHV 121  
 DB 283 IKF-----SFFPDWLSNVRSQTKPLFTVG-----EWSY----- 312  
 QY 122 RHFTGEGARIACOMQSIDELRHYQTETHAMSTYKFFNGFHHSNQWDRVWYLSVPKSEFF 181  
 DB 313 -----DINKLHNYNMTN--GTMSLFDAPLHN-----KFY 340  
 QY 182 EDAYSSGPFELTAVSFSEYVLTNLLFVPMGSAAYNGDMSTVTFGFSQAQSDSRHMTL 241  
 DB 341 TASKSGGTEDMRT-----LMTNTLM-----KDQPTL----- 366  
 QY 242 GIECIKFLLEQDPDNPVIVQWIDKWF-----WRGYRLL----- 275  
 DB 367 ---AVTFVDNHDTEPGALQSWVDWPKFLAYAFILTRQEGIPCVCYGYDYGIPQYNIPS 423

QY 276 -----TLVA-----MMMDYMPKRVMSW-RE-----SWEM 299  
 DB 424 LKSKIDPDLIARRDAYAYQTHDYLHSDILIGWTRGVIEKPGSLAALITDGPGRKWMY 483  
 QY 300 YAEQNGGALFCLDARYGIREPKGWODACEGKDHIHQAWSTFYGFNAASAFHTWYP--T 356  
 DB 484 VKQAGKGVFDLT--GNR-----SDTVT-----INSDGWGEF-KVNGGSV-SVWVPRKTT 530  
 QY 357 EDEMGLWSAKYP--DSFDYRYPFRDHW 382  
 DB 531 VSTIAWSITRPWTDEFVRWTEPRIVAW 558  
 RESULT 13  
 W49873  
 ID R04111 standard; Protein; 684 AA.  
 AC W49873;  
 DT 21-DEC-1998 (first entry)  
 DE Thermotoga maritima MSB8-6GB4 glycosidase.  
 KW Glycosidase; MSB8-6GB4; thermostable enzyme; oligosaccharide;  
 OS glucose; sugar; baking; textile; detergent; beta-galactosidase.  
 SW Thermotoga maritima strain MSB8-6GB4.  
 PN W09824799-A1.  
 PD 11-JUN-1998  
 PF 08-DEC-1997; U22623.  
 PR 10-OCT-1997; US-949026.  
 PA 06-DEC-1996; US-056916.  
 PI (DIVE-) DIVERSA CORP.  
 PI Bylina EJ, Lam DE, Mathur EJ, Swanson RV;  
 DR WPI: 98-362407/31.  
 DR N-PSDB: V36922.  
 PT Glycosidase enzymes from organisms of the genera Staphylococcus,  
 PT Pyrococcus and Thermococcus - for deriving sugar from  
 PT oligosaccharides, useful in the e.g. food processing, textile or  
 PT baking industries  
 PS Claim 1: Fig 16a-c: 92pp; English.  
 CC This is the amino acid sequence of glycosidase MSB6-6GB4, deduced  
 CC from a polynucleotide (see V36922) of clone 6GB4 of Thermotoga  
 CC maritima MSB8 that grows optimally at 85 degC in high salt medium.  
 CC The invention provides 18 polynucleotides (see V36907-24) coding  
 CC for thermostable glycosidases (see W49858-75) having glucosidase,  
 CC alpha-galactosidase, beta-galactosidase, beta-mannosidase,  
 CC beta-mannanase, endoglucanase or pullulanase activity. Vectors and  
 CC host cells are also claimed. A method is provided for producing  
 CC the enzymes by recombinant techniques. A claimed method for  
 CC generating glucose from soluble cell oligosaccharides comprises  
 CC contacting a sample (selected from dairy products, fruit juice,  
 CC detergent, textile, guar gum, animal feed, plant biomass or waste  
 CC product) containing oligosaccharides (selected from maltose,  
 CC cellobiose, lactose, sucrose, raffinose, stachyose, verbascose,  
 CC cellulose, starch, amylose, glycogen, disaccharides, polysaccharides  
 CC and pullulan) with one of the claimed glycosidases such that glucose  
 CC is produced.  
 SQ Sequence 684 AA;  
 Query Match 3.4%; Score 98.5; DB 1; Length 684;  
 Best Local Similarity 18.3%; Pred. No. 0.17;  
 Matches 95; Conservative 67; Mismatches 159; Indels 197; Gaps 29;  
 QY 54 DPFRLTMDAYWKY-----OGEKKLYAVIDFTQ---NNAFGLGVSDARYI--- 96  
 DB 49 DLFKIEDREWIYEREFKEDVKGERVDLVFEGVDTLSVDYLVNGVYLGSTEDMFIEVR 108  
 QY 97 -----NALKLFQ-----GV-----TPLEYLAHRGFAHVRHFTG-- 126  
 DB 109 FQVNVNLEKHNKLVKIPSPKRVKTLQSNQYGLVGGPEDI-----RGYIRKAQTSYGMW 163  
 QY 127 EGARIACQ-----MOSIDELRHQYQTHAMSTYNNKFFNGFHH----- 163  
 DB 164 WGARIVTSGIWKPVYLEYVYRARIQDSTAYLLEEGKDALVR-VNGFVHGEGLNIVEYVN 222  
 QY 164 -----SNQWFDRVWYLVSPKSFEDAYSSGPFPEFLTAVSFEYVLNLLFV 210

DB 223 GEKIGEPVLEKNGEKLFDGVFHLKDVKLWY-----PNNVGKPYLYDFVFLKDL--- 272  
 QY 211 PMSGAAAYNGDMSTVTFGSQAQSDSRHMTLGIETKFLLEQDPDNPV-IVORWIDKWFV 269  
 DB 273 -----NGEI-----YREEKK-----IGLRKRVIVOEPEDEGKTFFIENGKVFV 312  
 QY 270 RGYRLITLVAMMDYMPKRVMSW--RESWEMY-----AEON-----GGALFKDLARYG 316  
 DB 313 KG-----ANWIPSENILTWLKEEYKLVKMARSAANMLRVWGGGIYEREIFR 362  
 QY 317 IREPKG--WQD---AC-EGKDH-----SHQAWSTFYGFNAASAFHTWVPTEDENGW 362  
 DB 363 LCDELGLMVWQDFMVACLEYPDHLPWFRKLANEERAKIVRKLYHPSIVLWC-GNNENNW 421  
 QY 363 LSAKYPDSFDYRYPFRDHWGEQAR-----AGNREYMKTLPLMC--QVCQIPMLFTEP- 413  
 DB 422 -----GFDEWGNMARKVDGINLGNRLYLFDFPEICAEDPSTPYWPSSEY 466  
 QY 414 ----GNPTKIGARE-----SNLGNKFHFCSD 436  
 DB 467 GGEKANSEKGRHVVWYVWSGWMNYENYEKNYGRFISE 504  
 RESULT 14  
 R04111  
 ID R04111 standard; peptide; 1306 AA.  
 AC R04111;  
 DT 07-SEP-1990 (first entry)  
 DE Human angiotensin converting enzyme (ACE)  
 KW human angiotensin converting enzyme; hypertension; bradykinin.  
 OS synthetic.  
 FH Key  
 FT protein  
 FT Location/Qualifiers  
 FT 30..1277  
 FT /label-mature ACE  
 FT /note="derived from pre-ACE by removal of signal peptide"  
 FT modified\_site 38..38  
 FT /label-putative N-glycosylation site  
 FT modified\_site 54..56  
 FT /label-putative N-glycosylation site  
 FT modified\_site 74..76  
 FT /label-putative N-glycosylation site  
 FT modified\_site 111..113  
 FT /label-putative N-glycosylation site  
 FT modified\_site 146..148  
 FT /label-putative N-glycosylation site  
 FT modified\_site 160..162  
 FT /label-putative N-glycosylation site  
 FT modified\_site 318..320  
 FT /label-putative N-glycosylation site  
 FT modified\_site 445..447  
 FT /label-putative N-glycosylation site  
 FT modified\_site 509..511  
 FT /label-putative N-glycosylation site  
 FT modified\_site 523..525  
 FT /label-putative N-glycosylation site  
 FT modified\_site 677..679  
 FT /label-putative N-glycosylation site  
 FT modified\_site 713..715  
 FT /label-putative N-glycosylation site  
 FT modified\_site 760..762  
 FT /label-putative N-glycosylation site  
 FT modified\_site 942..944  
 FT /label-putative N-glycosylation site  
 FT modified\_site 1191..1193  
 FT /label-putative N-glycosylation site  
 FT modified\_site 1225..1227  
 FT /label-putative N-glycosylation site  
 FT modified\_site 1225..1227  
 FT /label-putative N-glycosylation site  
 PN W09003435-A.  
 PD 05-APR-1990.  
 PF 27-SEP-1989; F00496.  
 PR 27-SEP-1988; FR-012620.  
 PA (INRM) Institut National de la Sante et de la Recherche Medicale.

PI Soubrier F, Albenc-Gelas F, Hubert C, Corvol P;  
DR WPI: 90-132272/17.  
DR N-PSDB; Q04027.  
PT New DNA encoding human angiotensin converting enzyme used eg in  
PT diagnosis of hypertension, evaluation of enzyme inhibitors  
PS Disclosure; p; French.  
CC Human angiotensin converting enzyme hydrolyses angiotensin I and kinins.  
CC Either intact enzyme or fragments thereof can be used to generate  
CC antibodies for diagnostic use. Oligonucleotide probes can also be made  
CC which are complementary to the sequence encoding the enzyme.  
SQ Sequence 1306 AA;

Query Match 3.3%; Score 94.5; DB 1; Length 1306;  
Best Local Similarity 18.8%; Pred. No. 1.2;  
Matches 142; Conservative 80; Mismatches 214; Indels 321; Gaps 46;

QY 1 MDPTLKKLGLDRYAAMTGLGWETTY---QPMKDV--FPDYREGIKIHWDKW--- 52  
Db 428 VSTPEHLHKICLLDR---VTNDTESDINYLKMALEKIAFLPF---GYLVDQW-RWGVE 479

QY 53 ---VDPRLTWDAW---KYQEKELKYAVIDAFTONNAFLGVSDARYINALKLFQ 105  
Db 480 SGRTPPSRYNFD-WWYLRTKYQ-----ICPPVTRN-----ETHFDAGAKFHPN 523

QY 106 VTP-----LEYLAHRCFAHVGRH-----FTGEGARIACOMO----- 136  
Db 524 VTYRIRYVSFVLOFQHEALCKEAGYEGPLHQCIDIYRSYKAGAKLKVLOAGSSRPQWE 583

QY 137 -----SIDE---LRHYQTEPHAMSTYKFNFGHHNQFDRVWYLSVPKSFEDA 184  
Db 584 VLKDMVGLDALDAQPLLKYEPQVQWLQEQNO-QNG-EVLGWPEYQWHPLP-----DN 635

QY 185 YSSGPFEEFLT---AVFSFEYVLTNLLFPFMSGA-AYNGDMSTVT----- 226  
Db 636 YPEG-IDLVTDDEAFSEVEYDRTSQVWVNEAYEAANNYNTITITTSKILLQKNMQIA 694

QY 227 ---FGFSAQS-----DESRHMTLIGIE----- 244  
Db 695 NHTLKYGTQARKPDVNLQNTTIKRIIKVQDLERAAALPAQELLEYKNILLDMETYSVA 754

QY 245 -----CIKFLLEQPDNDVPIVORWIDK-WFWRGYRLTIVAMMDYMPKRVM--- 291  
Db 755 TVCHPNQSCILQ--LEPDLTNVMTSRKYEDLLWAEQWRDKAGRAILLQFY-PKYVELLN 810

QY 292 -----SWRESWENYA-EQNGGALFKDL-----ARYGIREPKGWQDAC 327  
Db 811 QARLNGYVDAGDSWRSMTYETPSLEQDLERLFOELQPLYLNLHAYVRRALHRYGAO--- 867

QY 328 EGKDHIS-----HQAWSFYG---FNAASAFHTWVPTED--EMGWSAKY 367  
Db 868 ---HINLEGPAPALLGNWQAOTWSNYDLVVFPPSAPSDT---TEAMLKQGTWTPRM 920

QY 368 PDSEDRYR-----PRDHWG-----EQARAGNR-----FY----- 393  
Db 921 FKEADDFETSLGLFVPPPEF--WNKSMLEKPTDGREVYCHASAMDFYNGDKFRKQCTTV 978

QY 394 -----MKTLPMICOTCOLPMLFTPEGNETKIGARES----- 424  
Db 979 NLEDLVVAHEMGHIQFMQKDLPAVLRGANP---GFHRAIGDVALSVSTPKHLHSL 1035

QY 425 NYLGNK-----PHFCSDHCK-DIFDHE---PORYVQAWLPVHQ 458  
Db 1036 NLLSSEGSDHEDINFLMKMALDKIAIFPSYLVQDWRWRVFDGSIITKENYQEWWSLRL 1095

QY 459 IHOCNCFPP-----DADPGAGEDPLAAYLDIYAVT 489  
Db 1096 KYQGLC-PPVPTQGDGPDGAKFIPSPVPIRYFVS 1131

RESULT 15  
W69165  
ID W69165 standard; Protein; 1965 AA.

AC W69165;  
DT 07-DEC-1998 (first entry)  
DE Streptococcus pneumoniae Iga protease.  
KW Iga protease; meningitis; pneumonia; antibiotic; vaccine.  
OS Streptococcus pneumoniae strain P110.  
FH Key Location/Qualifiers  
FT Region 96..132  
FT /note= "region resembles C-terminal anchor for  
FT other Gram-positive surface proteins"  
FT Region 96..101  
FT /note= "LPNTGS motif"  
FT Region 102..125  
FT /note= "hydrophobic region"  
FT Region 126..132  
FT /note= "lysine-rich charged region"  
FT Region 187..194  
FT /note= "ATP/GTP binding consensus"  
FT Region 396..628  
FT /note= "hydrophilic region comprising multiple  
FT tandem repeats"  
FT Peptide 396..415  
FT /note= "repeat 1"  
FT Peptide 416..436  
FT /note= "repeat 2"  
FT Peptide 437..453  
FT /note= "repeat 3"  
FT Peptide 454..470  
FT /note= "repeat 4"  
FT Peptide 472..483  
FT /note= "repeat 5"  
FT Peptide 485..501  
FT /note= "repeat 6"  
FT Peptide 507..523  
FT /note= "repeat 7"  
FT Peptide 534..546  
FT /note= "repeat 8"  
FT Peptide 547..560  
FT /note= "repeat 9"  
FT Peptide 595..607  
FT /note= "repeat 10"  
FT Peptide 614..628  
FT /note= "repeat 11"  
FT Binding\_site 1605..1609  
FT /note= "zinc binding domain (with residue 1629)"  
FT Cleavage\_site 1629  
FT /note= "zinc binding domain (with residues  
FT 1605-1609)"  
PN CA2196502-A.  
PD 24-MAR-1998.  
PR 31-JAN-1997; 196502.  
PR 23-SEP-1996; US-026939.  
PA (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.  
PA (NEW-) NEW ENGLAND MEDICAL CENT HOSPITALS INC.  
PI Gilbert-Rothstein JV, Plaut AG, Weiser JN;  
DR WPI: 98-399668/35.  
DR N-PSDB; V44592.  
PT New DNA encoding Streptococcus pneumoniae Iga protease - and  
PT antibiotic screening assay and subunit vaccine etc  
PS Claim 11; Fig 4; 66pp; English.  
CC This is the amino acid sequence of Streptococcus pneumoniae Iga  
CC protease, deduced from an isolated iga gene (see V69165). The  
CC invention provides a method of identifying a candidate antibiotic  
CC compound that involves incubating the Iga protease with an Iga  
CC protein, and comparing Iga protease activity in the presence and  
CC absence of a test compound. S. pneumoniae cells or isolated Iga  
CC protease enzyme can be used in this method. The antibiotic is  
CC used to treat S. pneumoniae or other bacterial infections. The  
CC invention also relates to a subunit vaccine comprising an isolated  
CC Iga protease polypeptide. This can be used to prevent or reduce  
CC the severity of (S. pneumoniae) infections.  
SQ Sequence 1965 AA;

Query Match 3.3%; Score 94; DB 1; Length 1965;  
Best Local Similarity 21.0%; Pred. No. 2.4;  
Matches 97; Conservative 56; Mismatches 193; Indels 116; Gaps 23;

QY 36 FPYDREYEGIK-----IHDWQKWDPERLTM-----DAYWKYQ-----CEKEKK 73  
Db 1222 FLYDRDSIVKEVLPDLQKLDYQSDAIRKTLGTSPEVKLTLEYLQCFQSKTKQNLGDSUKK 1281

QY 74 LYAVIDA--FTQNNAFLGVSADARYINALKLFLOGVTPLP-----YLAHRG 116  
Db 1282 LLSA-DAGLASHNSATRCYLYDKIKNNKEALLGLTYLERWYNENYGOVNVKDLVMYHPD 1340

QY 117 FAHVGR-----HFTGEGARIAQMGSID-----ELRHQYQETHAMST-----YNNKFENG 161  
Db 1341 FFRKGNTPDLTLIELGKSGFNLLAKNNVDYTGISLASQHGATDLFTLEHYRKVFLPN 1400

QY 162 HHSNQWF--DRVWVLSVPKSFEDA-----YSSGPEEFLTAVSFSEYVLTNLLF 209  
Db 1401 TSNDWFKSETKAYIVFEFKSTIEVKTQGLAGTKYSIGVYDRITSATWKYRNMVPLLT 1460

QY 210 VPMSGRAYNGDMSTVTGFSQAQSDSHMTLGIECIKFLLEQDPDNPVIVQRWIDKWF 269  
Db 1461 LPERSVFEVIS--TMSLGFAGDYRTSDHKA--GKALNDFVEENARETAKRQRDHYDW-- 1516

QY 270 RGYRLTLVAMMDYMQPKRYMSHRESWEMYAEQNGGALFKDLARYGIREPKGWODACEG 329  
Db 1517 --YRILV-----NSQPRKTLFVR--FSLY-----DAYKFGDDTTSG--KATAEA 1554

QY 330 KDHISHQAWSTIFYGFENASAFHTW---VPTDEMGWLSAKYPDSFDRY-YRPRFDH----- 381  
Db 1555 KFDSSNFAMKNFFGVPVGNKVYHNGHAYATGDBGVYMSYRMLDKHGAINYTHEMTHDSQ 1614

QY 382 -----WGEQARAGNRFRYMKTLPLMLCQTCQIPMLFTEPGNPT 417  
Db 1615 DIYLGCGRRNGLGPEFFAKGL-----LQAP---DQPSDPT 1647

This Page Blank (uspto)



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:22:40 ; Search time 60.65 seconds  
(without alignments)  
130,418 Million cell updates/sec

Title: US-09-430-029-5  
Perfect score: 2873  
Sequence: 1 MDPTTLKKLGLKDRYAAMT.....FDGSEDKNFAWRGQATRN 516

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 152396 seqs, 15329161 residues

Total number of hits satisfying chosen parameters: 152396

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pap.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pap.\*  
3: /cgn2\_6/ptodata/1/iaa/6\_COMB.pap.\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pap.\*  
5: /cgn2\_6/ptodata/1/iaa/backfiles.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2409	83.8	519	1	US-08-319-387-4
2	374.5	13.0	500	5	5171684-2
3	370	12.9	501	1	US-08-499-215-4
4	255	8.9	52	3	US-08-917-299-25
5	192	6.7	52	3	US-08-917-299-10
6	152	5.3	327	5	5171684-6
7	109	3.8	342	1	US-08-499-215-2
8	91.5	3.2	548	1	US-08-468-700-37
9	91.5	3.2	548	1	US-08-645-971-5
10	91.5	3.2	548	2	US-08-468-220-35
11	91.5	3.2	548	2	US-08-468-698-35
12	91.5	3.2	548	2	US-08-704-706A-37
13	91.5	3.2	548	3	US-08-890-383-6
14	91.5	3.2	548	3	US-08-914-679A-6
15	91.5	3.2	548	4	PCT-US94-01553A-35
16	91.5	3.2	548	4	PCT-US95-10426-35
17	90.5	3.2	336	3	US-09-095-163-2
18	90	3.1	217	2	US-08-176-414B-3
19	90	3.1	3200	2	US-08-477-451-8
20	89	3.1	868	2	US-08-162-081B-33
21	89	3.1	868	2	US-08-780-872-33
22	88	3.1	549	1	US-08-720-899-6
23	88	3.1	549	1	US-08-459-610-6
24	88	3.1	549	2	US-08-343-804-6
25	88	3.1	549	2	US-08-687-399-6
26	88	3.1	549	2	US-08-600-908A-6
27	88	3.1	549	3	US-08-683-838A-6
28	88	3.1	583	5	5256558-4

29 88 3.1 713 2 US-08-849-212-4 Sequence 4, Appli  
30 87 3.0 876 1 US-08-785-071A-2 Sequence 2, Appli  
31 87 3.0 876 3 US-09-012-872-2 Sequence 2, Appli  
32 86.5 3.0 972 3 US-08-335-844A-24 Sequence 24, Appli  
33 85.5 3.0 475 2 US-08-272-255-13 Sequence 13, Appli  
34 85.5 3.0 475 4 PCT-US95-08565-13 Sequence 13, Appli  
35 84.5 2.9 604 4 PCT-US96-12860-4 Sequence 4, Appli  
36 80.5 2.8 365 2 US-08-833-610-7 Sequence 7, Appli  
37 80.5 2.8 365 3 US-08-834-033A-17 Sequence 17, Appli  
38 80 2.8 52 3 US-08-917-299-26 Sequence 26, Appli  
39 80 2.8 976 3 US-08-750-141A-1 Sequence 1, Appli  
40 79.5 2.8 529 5 5217865-2 Patent No. 5217865  
41 79.5 2.8 604 2 US-08-511-485-6 Sequence 6, Appli  
42 79 2.7 866 1 US-08-386-727-8 Sequence 8, Appli  
43 79 2.7 866 2 US-08-600-452A-8 Sequence 8, Appli  
44 78.5 2.7 438 4 PCT-US95-05922A-2 Sequence 2, Appli  
45 78.5 2.7 455 2 US-08-392-625-21 Sequence 21, Appli

## ALIGNMENTS

RESULT 1  
US-08-319-387-4  
; Sequence 4, Application US/08319387  
; Patent No. 5543317  
; GENERAL INFORMATION:  
; APPLICANT: Shields, Malcolm S.  
; TITLE OF INVENTION: Microbial Degradation of Trichloroethylene,  
; TITLE OF INVENTION: Dichloroethylenes and Aromatic Pollutants  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/319,387  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/167,457  
; FILING DATE: 15-DEC-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/694,718  
; FILING DATE: 02-MAY-1991  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Whitlock, Ted W.  
; REGISTRATION NUMBER: 36,965  
; REFERENCE/DOCKET NUMBER: UWF-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 904-375-8100  
; TELEFAX: 904-372-5800  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 519 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-319-387-4



QY 79 DAFQONAFELGVSADYINAKLFLQGVTPLEYLEAHRGFAHVRHFTGEGARTACQMST 138  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-917-299-25  
Db 65 DGAVRTREATRI-EPFAEAMKIMVPLTNAEYQAVAGCGMILSAVENQELRGYRAQML 123  
QY 139 DELRHVOTETHAMSTYNKEFGHSHNQWDRWYLSVPKSF----- 180  
; LENGTH: 52 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-917-299-25  
Db 124 DEVRIHAQLE---MTLRN-----YAKHWC-----PSGFDIGQRGYQHYPAGLVSIG 167  
QY 181 -FEDAVSSGFFFLAVERSEFYVLTNLLFVPMSCAAVNGDMSTVTFGFSQSDSRHM 239  
; LENGTH: 52 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-917-299-25  
Db 168 EQHFNTGDPDVIIDNLVAETAFTNILLVATPQVAVANGDMASVFLSQSDDEARM 227  
QY 240 ILGIECIKFLLODDPNVPQVQIDKFWRGYRLT-TLVAMMDYMQPKRYMSRWSE 298  
; LENGTH: 52 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-917-299-25  
Db 228 ANGYGSVWALLENE-DNLPNQLQSLDRHFWRAHKALDNAVSGSCSEYGARKRPWSYKAQWE 286  
QY 299 MYAEON--GGALFKDLARYGIREPKGWODACEG-----KDHISHQAWSTFYGFNAAS 348  
; LENGTH: 52 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-917-299-25  
Db 287 EWVUDFVGGYIDR-LSEGVQAP-----ACLGAAADEVKVSHHTLGOVLSAVWPLN--- 337  
QY 349 AFHTWVPTDEMG-----WLSAKYPSDFRYRPRFDHW-GEQA-----RAGNRFYMKTLPM 399  
; LENGTH: 52 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-917-299-25  
Db 338 ---FW--RSDAMPADFEWFENHYPCWSAYQ-----GYWEGYKALADPAGGRIMLOELPG 388  
QY 400 L---CQTQOIPMLFTEGNTKIGARESNYLGNKFHSCDHCKDIPDHPQRY---VQAW 453  
; LENGTH: 52 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-917-299-25  
Db 389 LPPMCQVCQVPCVMP---RLDMNAARIIEFEGKIALCSEPQRIPTNWPAYRHRKQYW 445  
QY 454 LPVH 457  
Db 446 ARYH 449

## RESULT 4

US-08-917-299-25  
; Sequence 25, Application US/08917299  
; Patent No. 6010855  
; GENERAL INFORMATION:  
; APPLICANT: JACKSON, Mary  
; APPLICANT: GLOUEL, Brigitte  
; TITLE OF INVENTION: DESATURASE ANTIGEN OF MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: c/o Finnegan Henderson, Farrabow, Garrett &  
; ADDRESSEE: Dunner, L.L.P.  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/917,299  
; FILING DATE: 25-JUL-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/022,713  
; FILING DATE: 26-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MEYERS, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
; REFERENCE/DOCKET NUMBER: 03495.0156-00  
; TELEPHONE: (202) 408-4400  
; TELEFAX: (202) 408-4400  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:

QY 192 FLTAVSFSEFYVLTNLLFVPMSCAAVNGDMSTVTFGFSQSDSRHMTLGI 243  
; LENGTH: 52 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-917-299-25  
Db 1 FLTAVSFSEFYVLTNLLFVPMSCAAVNGDMSTVTFGFSQSDSRHMTLGL 52  
RESULT 5  
US-08-917-299-10  
; Sequence 10, Application US/08917299  
; Patent No. 6010855  
; GENERAL INFORMATION:  
; APPLICANT: JACKSON, Mary  
; APPLICANT: GLOUEL, Brigitte  
; TITLE OF INVENTION: DESATURASE ANTIGEN OF MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: c/o Finnegan Henderson, Farrabow, Garrett &  
; ADDRESSEE: Dunner, L.L.P.  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/917,299  
; FILING DATE: 25-JUL-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/022,713  
; FILING DATE: 26-JUL-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MEYERS, Kenneth J.  
; REGISTRATION NUMBER: 25,146  
; REFERENCE/DOCKET NUMBER: 03495.0156-00  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 408-4000  
; TELEFAX: (202) 408-4400  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 52 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-917-299-10  
Query Match 8.9%; Score 255; DB 3; Length 52;  
Best Local Similarity 94.2%; Pred. No. 1.7e-21;  
Matches 49; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 97 NALKFLQGVTPLEYLEAHRGFAHVRHFTGEGARTACQMSTIDELRHVQTE 147  
; LENGTH: 52 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-917-299-10  
Db 1 NALKFLTAVSPLEYPQAFQGSFVGRVSGAGARVACQMAIDELRHVQTO 51  
Query Match 6.7%; Score 192; DB 3; Length 52;  
Best Local Similarity 70.6%; Pred. No. 2e-14;  
Matches 36; Conservative 7; Mismatches 8; Indels 0; Gaps 0;



Query Match 3.2%; Score 91.5; DB 1; Length 548;  
Best Local Similarity 19.1%; Pred. No. 0.16;  
Matches 82; Conservative 49; Mismatches 136; Indels 16

REFERENCE/DOCKET NUMBER

REFERENCE/DOCKET NUMBER

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 742-7555  
; TELEFAX: (415) 742-7217  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 548 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-468-220-35

Query Match 3.2%; Score 91.5; DB 2; Length 548;  
Best Local Similarity 19.1%; Pred. No. 0.16;  
Matches 82; Conservative 49; Mismatches 136; Indels 163; Gaps 23;  
  
QY 21 RGLG--WETTYQPMKVPFPDYREGI-----KIHWDKW-----VDPFRLTMDAY 63  
DB 213 RGIKAWDEWDTENGNDYLMYADLMDHPVVTELKNGKVVNTNIDGFR-----267  
QY 64 WKYQGEKKLYAVIDAFQNNAPLGVSDARYINALKLFLOGVTPLEYLAHGFHVRH 123  
DB 268 ---DGLKHIFSEFPD-----WLSYVRSQTKPLFTVG-----EWSY-----302  
QY 124 FTGEGARIACQMSIDELRHVOTETHAMSTYKFNFGFHSNQWFDKRWYLSVPKSPFD 183  
DB 303 -----DINKLHNYITKN--GTMSLFADPLHN-----KFTYA 332  
QY 184 AYSGPPEFLTAVSFSEYVLTNLLFVPMGAAVNGDMST--VTFGSAQSDSRHMTL 241  
DB 333 SKSGGAFDMRT-----LMTNTLM-----KDQPTLAVTFVDNHDINPAKRCSH 374  
QY 242 GIECIK-----FLEQDPD-----NPIVQWIDKWF--WRGYRLITIVA 279  
DB 375 GRPWFKPLAYAFILTRQGYPCVFGYGYIPQYNIPSLKSIDPLLIARRDYAYGT---431  
QY 280 MMDYMQPKRVMSW-----RESWEMAEQNGGALFKDLARYGI 317  
DB 432 -QHDYLDHSDIIGTREGVTEKPGSGLAALITDGAGRSKMWYVKGQKAGKVFYDIT--GN 488  
QY 318 REPKGWQDACEGKHISHQAWSTFYGFNAASAFHTWVPTDEMGL-----SAKYPDSFD 372  
DB 489 R-----SDTVT-----INSDGWGEF-KVNGGSV-SWVVPKTTVSTIARPIITRPWTGEFV 537  
QY 373 RYRPRFDHW 382  
DB 538 RWHEPRLVAV 547

RESULT 11  
US-08-468-698-35  
; Sequence 35, Application US/08468698  
; Patent No. 5849549  
; GENERAL INFORMATION:  
; APPLICANT: Antrim, Richard L.  
; APPLICANT: Barnett, Christopher  
; APPLICANT: Mitchinson, Colin  
; APPLICANT: Power, Scott D.  
; APPLICANT: Requaadt, Carol  
; APPLICANT: Solheim, Leif P.  
; TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase  
; NUMBER OF SEQUENCES: 68  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genencor International, Inc.  
; STREET: 180 Kimball Way  
; CITY: South San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb  
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,698  
; FILING DATE: 06-JUN-95  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/194,664  
; FILING DATE: 10-FEB-94  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/016,395  
; FILING DATE: 11-FEB-93  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stone, Christopher L.  
; REGISTRATION NUMBER: 35,696  
; REFERENCE/DOCKET NUMBER: GC220D2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 742-7555  
; TELEFAX: (415) 742-7217  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 548 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-468-698-35

Query Match 3.2%; Score 91.5; DB 2; Length 548;  
Best Local Similarity 19.1%; Pred. No. 0.16;  
Matches 82; Conservative 49; Mismatches 136; Indels 163; Gaps 23;  
  
QY 21 RGLG--WETTYQPMKVPFPDYREGI-----KIHWDKW-----VDPFRLTMDAY 63  
DB 213 RGIKAWDEWDTENGNDYLMYADLMDHPVVTELKNGKVVNTNIDGFR-----267  
QY 64 WKYQGEKKLYAVIDAFQNNAPLGVSDARYINALKLFLOGVTPLEYLAHGFHVRH 123  
DB 268 ---DGLKHIFSEFPD-----WLSYVRSQTKPLFTVG-----EWSY-----302  
QY 124 FTGEGARIACQMSIDELRHVOTETHAMSTYKFNFGFHSNQWFDKRWYLSVPKSPFD 183  
DB 303 -----DINKLHNYITKN--GTMSLFADPLHN-----KFTYA 332  
QY 184 AYSGPPEFLTAVSFSEYVLTNLLFVPMGAAVNGDMST--VTFGSAQSDSRHMTL 241  
DB 333 SKSGGAFDMRT-----LMTNTLM-----KDQPTLAVTFVDNHDINPAKRCSH 374  
QY 242 GIECIK-----FLEQDPD-----NPIVQWIDKWF--WRGYRLITIVA 279  
DB 375 GRPWFKPLAYAFILTRQGYPCVFGYGYIPQYNIPSLKSIDPLLIARRDYAYGT---431  
QY 280 MMDYMQPKRVMSW-----RESWEMAEQNGGALFKDLARYGI 317  
DB 432 -QHDYLDHSDIIGTREGVTEKPGSGLAALITDGAGRSKMWYVKGQKAGKVFYDIT--GN 488  
QY 318 REPKGWQDACEGKHISHQAWSTFYGFNAASAFHTWVPTDEMGL-----SAKYPDSFD 372  
DB 489 R-----SDTVT-----INSDGWGEF-KVNGGSV-SWVVPKTTVSTIARPIITRPWTGEFV 537  
QY 373 RYRPRFDHW 382  
DB 538 RWHEPRLVAV 547  
  
RESULT 12  
US-08-704-706A-37  
; Sequence 37, Application US/08704706A  
; Patent No. 5958739  
; GENERAL INFORMATION:  
; APPLICANT: GENENCOR INTERNATIONAL, INC.  
; APPLICANT: COLIN MITCHINSON



QY 373 RYRPRFDHW 382  
Db 538 RWHEPRLVAW 547

## RESULT 14

US-08-914-679A-6  
; Sequence 6, Application US/08914679A

; Patent No. 6080568

; GENERAL INFORMATION:

; APPLICANT: Barbara A. Swanson

; TITLE OF INVENTION: MUTANT ALPHA-AMYLASE COMPRISING MODIFICATION

; TITLE OF INVENTION: AT RESIDUES CORRESPONDING TO A210, H405 AND/OR

; TITLE OF INVENTION: T412 IN BACILLUS LICHENIFORMIS

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genencor International, Inc.

; STREET: 925 Page Mill Road

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304-1013

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/914,679A

; FILING DATE: To Be Assigned

; ATTORNEY/AGENT INFORMATION:

; NAME: Stone, Christopher

; REGISTRATION NUMBER: 35,696

; REFERENCE/DOCKET NUMBER: GC387

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650) 846-7555

; TELEFAX: (650) 845-6504

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 548 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

; US-08-914-679A-6

Query Match 3.2%; Score 91.5; DB 3; Length 548;

Best Local Similarity 19.1%; Pred. No. 0.16;

Matches 82; Conservative 49; Mismatches 136; Indels 163; Gaps 23;

QY 21 RGLG--WETTYQPMKVPFYDRIYEGI-----KIHWDKW-----VDPRLTMDAY 63  
Db 213 RGIGKAWDEVDTEGNYDLYMADLMDHDPVTELNKNGKWKVYNTTNDGFR-----267  
QY 64 WKYQGEKELLYAVIDAFTONNAFLGVSDARYINALKFLQGVTPLEYLAHRGFVHGRH 123  
Db 268 ---DGLKHIFSPFD-----WLSYVRSQTGKPLFTVG-----EWSY-----302  
QY 124 FTGEGARIACQMSIDELRHYQVTHAMSYNKFNFHNSQWDFRWWVLSVPKSPFD 183  
Db 303 -----DINKLHNYITKTN--GTMSLFDAPLHN-----KEYTA 332  
QY 184 AYSGPPEFLTAVSFSEYVLTNLLFVPMGGAAYNGDMST--VTFGFSQSDSRHMTL 241  
Db 375 GRPWFKPLAVAFILTRQEGYPCVYGYGIPQYNIPSLKSKIDPLDIARRDYAIGT---431

QY 280 MMDYMQPKRVMWSW-----RESWEMYAEQNGGALFKDLARYGI 317  
Db 432 -QHDYLDHSDIICWTRGVTEKPGSGLAALLTDAGRSKMWYVKQKQAGKVFYDLT--GN 488  
QY 318 REPKGWQDACEGKDHISHQAWSTFYGFNAASAFHTWVPTDEMGWL-----SAKYDPDSFD 372  
Db 489 R-----SDTVT-----INSDGWGEF-KVNGGSV-SVWVPRKTTVTSTIARPTITRPTWGEFV 537  
QY 373 RYRPRFDHW 382  
Db 538 RWHEPRLVAW 547

## RESULT 15

PCT-US94-01553A-35

; Sequence 35, Application PCT/US9401553A

; GENERAL INFORMATION:

; APPLICANT: GENECOR INTERNATIONAL, INC.

; TITLE OF INVENTION: Oxidatively Stable Alpha-Amylase

; NUMBER OF SEQUENCES: 68

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genecor International, Inc.

; STREET: 180 Kimball Way

; CITY: South San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/01553A

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Horn, Margaret A.

; REGISTRATION NUMBER: 33,401

; REFERENCE/DOCKET NUMBER: GC220-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 742-7536

; TELEFAX: (415) 742-7217

; INFORMATION FOR SFO ID NO: 35:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 548 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; PCT-US94-01553A-35

Query Match 3.2%; Score 91.5; DB 4; Length 548;

Best Local Similarity 19.1%; Pred. No. 0.16;

Matches 82; Conservative 49; Mismatches 136; Indels 163; Gaps 23;

QY 21 RGLG--WETTYQPMKVPFYDRIYEGI-----KIHWDKW-----VDPRLTMDAY 63  
Db 213 RGIGKAWDEVDTEGNYDLYMADLMDHDPVTELNKNGKWKVYNTTNDGFR-----267  
QY 64 WKYQGEKELLYAVIDAFTONNAFLGVSDARYINALKFLQGVTPLEYLAHRGFVHGRH 123  
Db 268 ---DGLKHIFSPFD-----WLSYVRSQTGKPLFTVG-----EWSY-----302  
QY 124 FTGEGARIACQMSIDELRHYQVTHAMSYNKFNFHNSQWDFRWWVLSVPKSPFD 183  
Db 303 -----DINKLHNYITKTN--GTMSLFDAPLHN-----KEYTA 332  
QY 184 AYSGPPEFLTAVSFSEYVLTNLLFVPMGGAAYNGDMST--VTFGFSQSDSRHMTL 241  
Db 333 SKSGGAFDMRT-----LMTNTLM-----KDQPTLAVTFVDNHDINPAKRCSH 374





This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 26, 2000, 20:11:49 ; Search time 75.19 Seconds  
(without alignments)  
37.172 Million cell updates/sec

Title: US-09-430-029-6

Perfect score: 666

Sequence: 1 MAVTALPKYDFPVKDAVEK.....HKDLISPTGLDGLGASF 118

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	370	55.6	93	W06803	Toluene ortho-mono
2	341.5	51.3	119	W98974	Alcaligenes sp. pr
3	78.5	11.8	495	R41362	TS1. Trehalose sy
4	78.5	11.8	495	W37441	Trehalose-6-phosph
5	78.5	11.8	495	W37427	Yeast trehalose-6-
6	74.5	11.2	666	W64590	Human SRCR protein
7	74.5	11.2	1785	W64591	Human SRCR protein
8	72.5	10.9	495	R44923	GSL1 glucose sensi
9	72.5	10.9	495	W49026	Saccharomyces cere
10	71.5	10.7	273	W23277	Bordetella pertuss
11	66.5	10.0	345	R14408	Nuclear factor C/E
12	65	9.8	261	R63805	Polyhydroxyalkanoa
13	65	9.8	422	R99090	Human interleukin-
14	65	9.8	423	R92814	Human chromosome 1
15	65	9.8	491	W52826	Human interlucan
16	65	9.8	597	W52821	Human interlucan
17	65	9.8	834	W52820	Human interlucan
18	64	9.6	1274	W82953	Human PRCC-TFE3 co
19	63.5	9.5	524	W32096	Human ALP. New nuc
20	63	9.5	105	W74887	Miniature swine re
21	62.5	9.4	876	Y04883	Human secreted pro
22	62.5	9.4	886	Y04884	Mycobacterium spec
23	62.5	9.4	1450	W30751	Rat phospholipase-
24	62	9.3	622	R92702	Rice HY4 gene prod
25	61.5	9.2	596	W48785	Thyroid peroxidase
26	61.5	9.2	917	R64974	Human 5' EST seque
27	60.5	9.1	58	Y12746	Pol region gene pr
28	60.5	9.1	376	R78800	T. thermophilus ga
29	60.5	9.1	376	W8490	Thermus thermophil
30	60.5	9.1	392	W73765	M. tuberculosis an
31	60.5	9.1	392	W73655	M. tuberculosis an
32	60.5	9.1	552	W71636	Omega-cyclohexane
33	60.5	9.1	810	R04574	Derived amino acid

34	60	9.0	123	1	W78326	Fragment of human
35	60	9.0	467	1	W76426	Glucocorticoid ind
36	60	9.0	467	1	Y05284	EGF-like homologue
37	60	9.0	467	1	Y13346	Amino acid sequenc
38	59.5	8.9	304	1	W13405	Eupenicillium bref
39	59.5	8.9	524	1	W32091	Porcine retrovirus
40	59.5	8.9	681	1	R92701	Arabidopsis blue 1
41	59.5	8.9	810	1	R11255	Marine Ii-4 recept
42	59.5	8.9	810	1	W13498	Murine interleukin
43	59.5	8.9	810	1	W48313	Mouse interleukin-
44	59.5	8.9	810	1	W60669	Mouse interleukin-
45	59.5	8.9	810	1	W73471	Mouse interleukin-

#### ALIGNMENTS

RESULT 1

W06803

ID W06803 standard; Protein; 93 AA.

AC W06803:

DE 29-JAN-1997 (first entry)

DE Toluene ortho-monooxygenase subunit tomA4.

KW tom; PTOM; self-transmissible; constitutive; bioreactor; pollutant;

KW breakdown; trichloroethylene; TCE; degradation.

OS Pseudomonas cepacia strain PRL-23.

PN US543317-A.

PD 06-AUG-1996.

PF 02-MAY-1991; 694718.

PR 02-MAY-1991; US-694718.

PR 15-DEC-1993; US-167457.

PR 06-OCT-1994; US-319387.

PA (FRAN//) FRANCESCONI S C.

PA (SHE//) SHIELDS M S.

PI Francesconi SC, Shields MS;

PT WPI; 96-370640/37.

PT Microorganisms transformed with P. cepacia PRL-23 Tom enzyme gene - are useful for degradation of chloro:aliphatic cpds. and aromatics

PS Claim 1; Column 33-34; 25pp; English.

CC The present sequence is that of toluene ortho-monooxygenase subunit tomA4 encoded by T44457, isolated from Pseudomonas cepacia strain PRL-23. The Tom gene is present on a large self-transmissible plasmid

CC denoted PTOM. The enzyme is capable of degrading trichloroethylene (TCE), a hazardous pollutant. The PTOM plasmid is transmissible and

CC expressable in other bacteria, thus many bacteria can be genetically altered to constitutively degrade TCE, esp. in bioreactors or

CC TCE-contaminated environments. P. cepacia PRL-23 contg. PTOM does not need exogenous chemical inducers and is capable of functioning under a

CC diverse set of conditions. Also it does not require an inducer that is a co-substrate for the enzyme required to break down TCE, in effect, it is

CC not subject to competitive inhibition.

CC Sequence 93 AA;

Query Match

Best Local Similarity 55.6%; Score 370; DB 1; Length 93;

Matches 63; Conservative 12; Mismatches 18; Indels 0; Gaps 0;

QY 1 MAVTALPKYDFPVKDAVEKFFAPILYVCWENHLMFPAPFCLPLPDMFPGALAGDVLPPV 60

Db 1 MSVTALPKYKFPARDARENFAPILFLIGWEDHLLFAAPVALPLPSDTILFGALCTQVLPST 60

QY 61 YGYHPDFAKIDWDVNRFSRGPWAPDPAKSLA 93

Db 61 YGYHPDFSKIDWSQVQWFKSQGPWHPDPAKSLA 93

RESULT 2

W98974

ID W98974 standard; Protein; 119 AA.

AC W98974;

DT 10-MAY-1999 (first entry)

DE Alcaligenes sp. protein Foxe.



PT by specific ribozyme or stimulator of trehalose-6-phosphate  
 PT production, particularly for in vitro or in vivo insulin production  
 PS Claim 13; Page 152-154; 263pp; English.  
 CC A mammalian cell has been developed comprising an effective amount of a  
 CC low Km hexokinase-inhibitor selected from: (a) an agent that stimulates  
 CC production of trehalose-6-phosphate (T6P); or (b) a ribozyme specific  
 CC for low Km hexokinase. The present sequence represents yeast  
 CC trehalose-6-phosphate synthase. The cells are particularly used to  
 CC produce insulin, in response to glucose or other secretagogues, either  
 CC in vitro or in vivo (for treating diabetes), but may also be used to  
 CC produce many other therapeutic proteins, e.g., from a gene under control  
 CC of the insulin promoter which is therefore responsive to glucose.  
 CC Growth-inhibited cells (having altered hexokinase activity) are used to  
 CC produce proteins, e.g., insulin or antibodies, in vivo or in vitro.  
 CC Reduction in low Km hexokinase activity provides cells in which insulin  
 CC secretion is induced at glucose concentrations closer to the normal  
 CC range than in the parent cell (nearly homeostatic secretion). Implanted  
 CC cells of reduced low Km hexokinase activity are expected to survive  
 CC longer in the host.  
 SQ Sequence 495 AA;

Query Match 11.8%; Score 78.5; DB 1; Length 495;  
 Best Local Similarity 24.7%; Pred. No. 0.59; Mismatches 23; Gaps 4;  
 Matches 23; Conservative 19; Mismatches 28; Indels 23; Gaps 4;  
 QY 14 KDAAKFFAPLLVGVWENHLMFPAPFCLPLPDPMPFAGLVPPVGYGHPDFAKIDWD 73  
 DB 80 KDLLEKFNAPFLSDE-----IADLYNGFSNLIWPLPHYHP-GEINFD 124  
 QY 74 RVWEF---RSGEWAPDPAKSLAGNGLGHKDLI 103  
 DB 125 ENAWFGYNEANQTFNTEAKT-----MNHNDLI 152

RESULT 6  
 W64590  
 ID W64590 standard; Protein; 666 AA.  
 AC W64590;  
 DT 23-OCT-1998 (first entry)  
 DE Human SRCR protein fragment.  
 KW Scavenger receptor cysteine rich domain; SRCR; diagnosis; treatment;  
 KW nervous system; medullo-blastoma; glioma; breast; detection;  
 KW autoantibody.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Misc\_difference 3  
 FT /label= unknown  
 PN WO9830687-A2.  
 PD 16-JUL-1998.  
 PF 09-JAN-1998; D00096.  
 PR 18-JUL-1997; DE-030997.  
 PR 09-JAN-1997; DE-000519.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
 PI Mollenhauer J, Poustka A;  
 DR WPI: 98-399136/34.  
 DR N-PSDB; V49651.  
 DT Proteins containing scavenger receptor, cysteine rich domain -  
 PT useful for diagnosis and treatment of tumours  
 PS Claim 1; Fig 1; 54pp; German.  
 CC This sequence represents a fragment of a human protein which contains a  
 CC SRCR (scavenger receptor, cysteine-rich) domain. The gene and encoded  
 CC protein can be used to diagnose or treat tumours, particularly of the  
 CC nervous system (medullo-blastoma or glioma) or breast. The DNA sequence  
 CC and probes derived from it, are used to identify genes that express  
 CC SRCR-domain containing proteins, to determine the form in which these  
 CC proteins exist and to assess the significance of individual forms on  
 CC cellular properties. The protein can be used to detect the presence of  
 CC autoantibodies, and Ab which regulate its expression.  
 SQ Sequence 666 AA;

Query Match 11.2%; Score 74.5; DB 1; Length 666;

Best Local Similarity 38.9%; Pred. No. 2.3;  
 Matches 21; Conservative 6; Mismatches 20; Indels 7; Gaps 3;

QY 27 VCVENHLMFPAPFCLPLPDPMPF---GAL---AGDVLPVY-GYHPDFAKIDWD 73  
 DB 405 ICSGNHLSTPAPFLNITRPNTDYSQGFSLQSPGDFSSPFGNYPNNAKCVD 458

RESULT 7  
 W64591  
 ID W64591 standard; Protein; 1785 AA.  
 AC W64591;  
 DT 23-OCT-1998 (first entry)  
 DE Human SRCR protein.  
 KW Scavenger receptor cysteine rich domain; SRCR; diagnosis; treatment;  
 KW nervous system; medullo-blastoma; glioma; breast; detection;  
 KW autoantibody; ss.  
 OS Homo sapiens.  
 PN WO9830687-A2.  
 PD 16-JUL-1998.  
 PF 09-JAN-1998; D00096.  
 PR 18-JUL-1997; DE-030997.  
 PR 09-JAN-1997; DE-000519.  
 PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.  
 PI Mollenhauer J, Poustka A;  
 DR WPI: 98-399136/34.  
 DR N-PSDB; V49652.  
 DT Proteins containing scavenger receptor, cysteine rich domain -  
 PT useful for diagnosis and treatment of tumours  
 PS Claim 2; Fig 2; 54pp; German.  
 CC This sequence represents a human protein which contains a SRCR (scavenger  
 CC receptor, cysteine-rich) domain. The gene and encoded protein can be used  
 CC to diagnose or treat tumours, particularly of the nervous system  
 CC (medullo-blastoma or glioma) or breast. The DNA sequence and probes  
 CC derived from it, are used to identify genes that express SRCR-domain  
 CC containing proteins, to determine the form in which these proteins exist  
 CC and to assess the significance of individual forms on cellular  
 CC properties. The protein can be used to detect the presence of  
 CC autoantibodies and antibodies which regulate its expression.  
 SQ Sequence 1785 AA;

Query Match 11.2%; Score 74.5; DB 1; Length 1785;  
 Best Local Similarity 38.9%; Pred. No. 7.5; Mismatches 7; Gaps 3;  
 Matches 21; Conservative 6; Mismatches 20; Indels 7; Gaps 3;

QY 27 VCVENHLMFPAPFCLPLPDPMPF---GAL---AGDVLPVY-GYHPDFAKIDWD 73  
 DB 1356 ICSGNHLSTPAPFLNITRPNTDYSQGFSLQSPGDFSSPFGNYPNNAKCVD 1409

RESULT 8  
 R44923  
 ID R44923 standard; Protein; 495 AA.  
 AC R44923;  
 DT 05-JUL-1994 (first entry)  
 DE GGS1 glucose sensing protein.  
 KW Glucose; sensing; cell signalling; yeast; Saccharomyces cerevisiae;  
 KW bread; breadmaking; fermentation; alcohol; beverage; brewing;  
 KW drying; osmotic shock; freezing; sugar metabolism.  
 OS Saccharomyces cerevisiae.  
 PN EP-577915-A.  
 PD 12-JAN-1994.  
 PF 09-JUL-1992; 870102.  
 PR 09-JUL-1992; EP-870102.  
 PA (ALGI-) ALGIST-BRUGGEMAN NV.  
 PI Holmann S, Thevelein J, Van Dijck P;  
 DR WPI: 94-009855/02.  
 DR N-PSDB; Q54385.  
 DT Transformed yeast with increased stress resistance or  
 PT fermentation capacity - has modification in general glucose  
 PT sensor system, partic. for bread-making, but also prodn. of  
 PT alcohol or foreign proteins

PS Disclosure: Figure 1; 26pp; French.  
 CC Yeast strains transformed with an altered GGS1 gene are resistant to  
 CC stress and/or have an altered sugar metabolism. The transformed  
 CC yeast can be used in industry for the production of bread, alcohol  
 CC and beverages and has resistance to at least one of drying, osmotic  
 CC shock (especially in sugar containing doughs) and freezing.  
 CC Preferably the yeast also has higher trehalose content with delayed  
 CC trehalose mobilisation.  
 SQ Sequence 495 AA;

Query Match 10.9%; Score 72.5; DB 1; Length 495;  
 Best Local Similarity 23.7%; Pred. No. 2.7;  
 Matches 22; Conservative 19; Mismatches 29; Indels 23; Gaps 4;  
 QY 14 KDAVEKFPAPLLYVCWENHLMFPAPCLPLPDPMPFGALAGDVLPVYGYHDPFAKIDWD 73  
 DB 80 KDLLEKFNAPVPLSDE-----IADLHYNGFSNLSILWPLFHYHP--GEINFD 124  
 QY 74 RVEWFP---RSGEWPAPDPAKSIAGNGLGHKDLI 103  
 DB 125 ENAWLAYNEANQTFNTEIAKT-----MHNNDLI 152

RESULT 9  
 W49026  
 ID W49026 standard; protein; 495 AA.  
 AC W49026;  
 DT 07-OCT-1998 (first entry)  
 DE Saccharomyces cerevisiae trehalose synthase.  
 KW Trehalose synthase; stabilising agent; prokaryotic cell preservation;  
 KW prophylaxis; live bacterial vaccine; vaginal infection;  
 KW urinary tract infection; live bacterial neutraceutical.  
 OS Saccharomyces cerevisiae.  
 PN WO9824882-A1.  
 PD 11-JUN-1998.  
 PF 05-DEC-1997; G03375.  
 PR 05-DEC-1996; US-032423.  
 FA (QUAD-) QUADRANT HOLDINGS CAMBRIDGE LTD.  
 PI Colaco C, Dhaliwal KS, Roser BJ, Tunnaccliffe AG,  
 PI Welsh DT;  
 DR WPI: 98-33305/29.  
 PT Preservation of prokaryotic cells - by increasing intracellular  
 PT trehalose concentration, mixing with a stabilising agent, and drying  
 PT to produce a glass form of the stabilising agent

PS Disclosure: Pages 39-40; 64pp; English.  
 CC The present sequence represents the Saccharomyces cerevisiae trehalose  
 CC synthase. The invention provides methods of preserving prokaryotic  
 CC cells which involves inducing the intracellular trehalose production  
 CC of the prokaryotic cell to an amount which effectively allows increased  
 CC storage stability and drying of the cells in the presence of a  
 CC stabilising agent. Therefore, prokaryotic cells known to synthesise  
 CC trehalose, such as S. cerevisiae, can be used to illustrate the methods  
 CC of the invention. Bacteria stabilised by these methods showed less than  
 CC 10% loss of viability on storage even after being stored at temperatures  
 CC up to at least 37 degrees centigrade for as long as 6 weeks. The  
 CC invention claims the methods to be useful for producing dried, stable  
 CC prokaryotic cells such as bacteria useful for pharmacological treatment,  
 CC prophylaxis, agricultural and industrial applications. The invention  
 CC also claims the methods to be useful for producing live bacterial  
 CC vaccines in a dry stable form, live bacterial neutraceuticals in a dry  
 CC stable form and other live bacterial pharmaceuticals in a dry stable  
 CC form, e.g. for treatment of vaginal or urinary tract infections.  
 SQ Sequence 495 AA;

Query Match 10.9%; Score 72.5; DB 1; Length 495;  
 Best Local Similarity 23.7%; Pred. No. 2.7;  
 Matches 22; Conservative 19; Mismatches 29; Indels 23; Gaps 4;  
 QY 14 KDAVEKFPAPLLYVCWENHLMFPAPCLPLPDPMPFGALAGDVLPVYGYHDPFAKIDWD 73  
 DB 80 KDLLEKFNAPVPLSDE-----IADLHYNGFSNLSILWPLFHYHP--GEINFD 124

QY 74 RVEWFP---RSGEWPAPDPAKSIAGNGLGHKDLI 103  
 DB 125 ENAWLAYNEANQTFNTEIAKT-----MHNNDLI 152  
 RESULT 10  
 W23277  
 ID W23277 standard; Protein; 273 AA.  
 AC W23277; 1997 (first entry)  
 DT 03-NOV-1997 (first entry)  
 DE Bordetella pertussis ptSE.  
 KW Holotoxin; ptx; pts; gene; extracellular export; ptSE; secretion;  
 KW whooping cough; vaccine.  
 OS Bordetella pertussis.  
 PN US5643747-A.  
 PD 01-JUL-1997.  
 PF 15-MAR-1993; 031619.  
 PR 31-MAR-1994; US-221750.  
 PR 15-MAR-1993; US-031619.  
 PA (AMCY ) AMERICAN CYANAMID CO.  
 PI Baker SM, Deich RA;  
 DR WPI: 97-350237/32.  
 N-PSDB; T84745.  
 DR Production of secreted pertussis holotoxin - by culturing  
 PT microorganisms transformed with Bordetella pertussis ptx operon and  
 PT new pts secretion sequences  
 PS Claim 1; Columns 45-48; 3pp; English.  
 CC The present sequence is the Bordetella pertussis ptSE, which is  
 CC 1 of the proteins encoded by the pts gene that facilitate the  
 CC extracellular export of the pertussis holotoxin (ptx), namely  
 CC ptSAB, ptSC, ptSE, ptSF, and ptSG. A secreted ptx can be  
 CC produced by culturing an E. coli or Bordetella host transformed  
 CC with 1 or more vectors, comprising the ptx and pts genes  
 CC operatively linked to appropriate expression control sequences. The  
 CC secreted ptx can be used in the production of whooping cough  
 CC vaccines.  
 SQ Sequence 273 AA;

Query Match 10.7%; Score 71.5; DB 1; Length 273;  
 Best Local Similarity 21.6%; Pred. No. 1.8;  
 Matches 38; Conservative 20; Mismatches 47; Indels 71; Gaps 8;  
 QY 3 VIALKPYDFPKDAVEKFPAPLLYVCWENHLM-----WNTNLLVKTDRLYDFDLVLASADAATPQALORS 139  
 DB 83 VLYIKAKSFPAQSPQAQPEPGL---WNTNLLVKTDRLYDFDLVLASADAATPQALORS 139  
 QY 35 -----FP-----APFCLPLPDPMPFGAL-----AGDVLPVYGYHDPFAKI 70  
 DB 140 RMAYRLQFRYPAPQAASRAASPVGPAPAGALNRRYAMOVGVGSDGIAPTAAY-----192  
 QY 71 DMDRVEW---FRSGEPW-----APDPAKSIAGNGLGHKDLISFR--TPGLDGLGGAS 117  
 DB 193 DDGRHTWLTFRGPQPPFAVFAVAPDGTETLVNLHIDNQSILVHRVAPVLMRSGAS 248

RESULT 11  
 R14408  
 ID R14408 standard; Protein; 345 AA.  
 AC R14408;  
 DT 11-FEB-1992 (first entry)  
 DE Nuclear factor C/EBP2.  
 KW Interleukin 6; IL-6; transcription; hematinic.  
 OS Homo sapiens.  
 PN J03236782-A.  
 PD 22-OCT-1991.  
 PF 17-OCT-1990.  
 PR 25-DEC-1989; JP-336868.  
 PR 17-OCT-1990; JP-279650.  
 PA (CHUZ-) CHUZO KISH IMOTO.  
 DR WPI: 91-35113/48.  
 DR N-PSDB; Q15015.







---

**This Page Blank (uspto)**